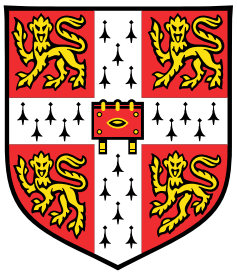
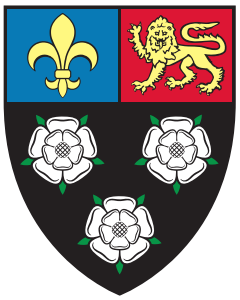


Three Essays in Financial Econometrics



Shuyi Ge

Faculty of Economics
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

King's College July 2021

I would like to dedicate this thesis to my loving parents and my beloved husband.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Shuyi Ge
July 2021

Abstract

Three Essays in Financial Econometrics

Shuyi Ge
King's College
University of Cambridge

Understanding how cross-sectional units interact with each other in a panel setting is an important question, given we are living in a more and more interconnected world. The effort to provide a solution to this question involves proposing statistical models that capture such features and obtain network datasets that characterize interdependency among entities. With a hope to contribute to this discipline, this thesis looks into the cross-sectional dependence in panel both theoretically and empirically. The first chapter develops a multi-country contagion model where the individual-specific Markov chains are interdependent. The second chapter studies a spatial factor model, which accommodates two distinct types of cross-sectional dependence in a panel. The chapter also utilizes a novel network dataset and empirically shows local interactions play a vital role in explaining comovement in equity returns. Chapter 3 studies peer groups of arbitrage characteristics. The details of the three chapters are summarized below:

Sovereign Risk Contagion in Eurozone with Mutual Exciting Regime-Switching Model

This paper proposes a new mutual exciting regime-switching model where crises can spread contagiously across countries. Each country has its own hidden stochastic process that determines whether the country is in a normal or crisis regime. Contagion is defined as a rise in the transition probability to the crisis regime when other countries are in crisis in the past state. Using this new approach, I revisit the sovereign risk contagion in the euro area. I

find that there are striking shifts in market pricing functions for the sovereign bond spreads. Multi-country contagion plays a dominant role in driving such shifts, while common risk factors and country-specific fundamentals are much less important.

News-Implied Linkages and Local Dependency in the Equity Market

This paper studies a heterogeneous coefficient spatial factor model that separately addresses both common factor risks (strong cross-sectional dependence) and local dependency (weak cross-sectional dependence) in the equity returns. For a high-dimensional panel of equity returns, it is challenging to measure firm-to-firm connectivity. We use extensive business news to construct firms' links via which local shocks transmit, and we use those news-implied linkages as a proxy for the connectivity among firms. We document a considerable degree of local dependency among *S&P* 500 stocks. From the asset pricing perspective, we derive the theoretical implications of no asymptotic arbitrage for the heterogeneous spatial factor model. Empirically, we show that adding spatial interactions to factor models significantly reduces mispricing and estimation errors. We also show that our news-implied linkages provide a comprehensive and integrated proxy for firm-to-firm connectivity, and it out-performs other existing networks in the literature.

Dynamic Peer Groups of Arbitrage Characteristics

This chapter proposes an asset pricing factor model constructed with semi-parametric characteristics-based mispricing and factor loading functions. We approximate the unknown functions by B-splines sieve where the number of B-splines coefficients is diverging. We estimate this model and test the existence of the mispricing function by a power enhanced hypothesis test. The enhanced test solves the low power problem caused by diverging B-splines coefficients, with the strengthened power approaches to one asymptotically. We also investigate the structure of mispricing components through Hierarchical K-means Clusterings. We apply our methodology to CRSP (Center for Research in Security Prices) and Compustat data for the US stock market with one-year rolling windows during 1967-2017. This empirical study shows the presence of mispricing functions in certain time blocks. We also find that distinct clusters of the same characteristics lead to similar arbitrage returns, forming a "peer group" of arbitrage characteristics.

Acknowledgements

This thesis owes much to many people. I would first like to thank supervisor Oliver Linton for his enormous help throughout this journey. I learn so much from him, not only how to be a qualified researcher but also the dedication to work and career. His patient guidance and warm encouragement have always been critical for me to complete this thesis. I am also indebted to Hashem Pesaran and Cheng Hsiao for their hospitality when I visited the University of Southern California in 2020. I am grateful to Hashem Pesaran for always sparing his time to read my work and giving me constructive comments. I am hugely inspired by his working attitude, time management skills, and dedication to students. I would like to thank Cheng Hsiao for being supportive and understanding. My thanks also go to Alexey Onatskiy for his inspiring conversations. I also thank Donald Robertson, Christopher Harris, and Koen Jochmans for being great course coordinators. I benefit so much from doing teaching assistant jobs for them.

I would like to thank the Faculty of Economics, Econometrics research group. I was kindly given a chance to present all my works in the thesis. I benefit a lot from people's attendance and comments. I particularly need to thank Shaoran Li, Weiguang Liu, Nicolas Paez, Rafe Martyn, Marco Valerio, Seok Young Hong, Merrick Li, Chen Wang for all those inspiring discussions.

I am most grateful to my beloved family, in particular, to my parents and my husband, to whom I dedicate this thesis. I need to thank them for always being supportive and having trust in me.

My thanks also go to Sophie Song, who has been an elderly sister to me in Cambridge. When I was distressed and unsure about my future, her kind support cured me.

I also thank all my friends in Cambridge, with whom I have had so much joyful time and shared many good memories. I thank Cambridge University Oriental Dance Association, King's College Badminton Socitey, Wolfson College Badminton Society, Cambridge Taek-

wondo Society, and Cambridge KPOP Society for bringing so much joy to my life in this beautiful city.

Chapter 1 is my solo work. Chapter 2 is joint work with Oliver Linton. Chapter 3 is joint work with Shaoran Li and Oliver Linton, which has been conditionally accepted by *Journal of Business and Economic Statistics*.

Contents

Chapter 1

Sovereign Risk Contagion in Eurozone with Mutual Exciting Regime-Switching Model

Abstract

This paper proposes a new mutual exciting regime-switching model where crises can spread contagiously across countries. Each country has its own hidden stochastic process that determines whether the country is in a normal or crisis regime. Contagion is defined as a rise in the transition probability to the crisis regime when other countries are in crisis in the past state. Using this new approach, I revisit the sovereign risk contagion in the euro area. I find that there are striking shifts in market pricing functions for the sovereign bond spreads. Multi-country contagion plays a dominant role in driving such shifts, while common risk factors and country-specific fundamentals are much less important.

Keywords: Contagion; Inter-dependence; Regime-switching; Mutual excitation; Sovereign credit risk
JEL Classification: C14; G11; G12

1.1 Introduction

The unfolding of the European Sovereign Debt Crisis shows that extreme events in the financial markets appear in clusters instead of in isolation. And this triggers a surge of interest in the role of contagion in the risk clustering. Testing for the existence of contagion and quantifying the size of it is important for economists and policymakers. This paper proposes a new mutual exciting regime-switching model where crises can spread contagiously across countries.

The first challenge to study contagion regards its definition. Asset returns exhibit co-movement due to exposure to common factors and spillovers. Is contagion just the normal time inter-dependence? Or does it reflect an increase in the co-movement during periods of crisis? This paper does not aim to contribute to this theoretical debate. Rather, I adopt the theoretical framework from the seminal paper of Forbes and Rigobon (2002), which defines contagion as the significant increase in the co-movement beyond what can be explained by normal time interactions. In particular, inter-dependence and contagion are distinguished. While inter-dependence is a result of normal market linkages, contagion is a breakdown of the normal time transmission regime.

Our second challenge is the specification of the crisis regime. Many studies implicitly assume that the crisis regime can be known *a priori*. For example, the correlation-based test for contagion (Boyer et al. (1997), Rigobon (2003), Corsetti et al. (2005)), the factor-model-based test for contagion (Dungey et al. (2002), Dungey et al. (2006), Dungey and Martin (2004), Bekaert and Harvey (2003)), and the structural break test for contagion (Bekaert et al. (2014), Beirne and Fratzscher (2013)) all rely on *a priori* identification of the crisis regime. These methods, while being very popular, given they are simple to implement and interpret, suffer from several problems. First of all, they implicitly assume that the crisis regime is continuous, which is not true according to the empirical findings in this paper. In addition, the *ex-post* nature of these methods makes them not particularly useful for detecting early-warning signals. Another strand of literature assumes the crisis regime is associated with some extreme values of the observed dependent variable. For example, Pesaran and Pick (2007) and Metiu (2012) assumes a country is in crisis regime when its endogenous performance variable is above a pre-specified threshold value. Caporin et al. (2018), in a similar vein, associates crisis regime with some high quantiles of the observed dependent variable. However, crises, in many cases, are more complicated than extreme values of a performance indicator. And they tend to be unobserved processes governed by several mechanisms.

The third challenge we face is the dynamics of the regime-switching process. There could be many reasons for a switch from the normal to crisis regime. In particular, we are interested in the role of contagion in the regime-switching process. That is, for one country, whether the transition to the crisis regime is more likely when other countries are in the crisis regime in the last period (i.e., crises spread contagiously across countries)? Moreover, how to quantify the strength of contagion, if it exists?

This paper proposes a new mutual exciting regime-switching model where crises can spread contagiously across countries. Each country has its own hidden stochastic process that determines whether the country is in a normal or crisis state. The country-specific process is governed by some common macroeconomic factors, country-specific fundamentals, and other countries' past states. Contagion is defined as a rise in the transition probability to the crisis regime when other countries are in crisis in the past state, after controlling for other mechanisms that drive the switching process. Inter-dependence and contagion are distinguished. There are three avenues for inter-dependence in the model. Firstly, asset returns are exposed to common factors and spillovers from others. Secondly, innovations to asset returns are allowed to be cross-sectionally correlated. Lastly, countries' hidden stochastic processes are subject to common risk factors.

The key element in the model that captures the multi-country contagion is the mutual exciting regime-switching process. The idea is closely related to some recent papers by Ait-Sahalia et al. (2014) and Aït-Sahalia et al. (2015). To capture the propagation of jumps across markets, those authors model the jump intensity using the mutual exciting jump process, also known as Hawkes process (Hawkes (1971b), Hawkes (1971a)) where the jump in one market could increase the probability of future jumps elsewhere. In their spirits, I model multi-country contagion using a mutual exciting regime-switching process. Under the framework, one country being in the crisis regime could increase the transition probability to the crisis regime for other countries. As in Ait-Sahalia et al. (2014) and Aït-Sahalia et al. (2015), the cross country contagion is probabilistic rather than certain. And the model goes beyond their work by allowing for a richer contagion pattern. In addition to a break in the mean equation, the model can also accommodate a break in the variance. The dynamics in the variance could be especially important for crisis episodes where increases in volatility are the major symptomatic.

Using the new approach, I revisit the sovereign risk contagion in the euro area. I use daily 10-year sovereign bond spreads of six euro area countries, including Greece, Ireland, Portugal, Spain, Italy (GIPSI countries), and France from 12/02/2008 to 01/12/2011. German government bond yields of the same maturity are used as the benchmark. I deliberately end

the sample before the European Central Bank (ECB) announced the Long-Term Refinancing Operations (LTRO) to avoid the clustering of switchings due to the intervention. Some interesting empirical results are found. Firstly, sovereign bond spread pricing functions are highly regime-dependent as there are striking shifts in market pricing behaviors. In the crisis regime, most sample countries experience a significantly positive jump in the intercept. There is a break of exposures to common risk factors. And the directions of the shifts are opposite to the sign of exposures in the normal regime. This might because the risk aversion and uncertainty both start falling since Spring 2009 while the euro area sovereign spreads begin to skyrocket after 2010. It suggests that the factor cannot help to interpret the sharp increases in euro area sovereign bond spreads during the European debt crisis. On the other hand, regional risk spillovers explain more variations in the sovereign bond spreads during periods of crisis as the vector auto-regressive coefficients are much larger in magnitude in the crisis regime. Surprisingly, Greece plays a less important role in directly propagating shocks to others. This is because investors start to isolate Greek bonds from other countries when the Greek default is inevitable. As a result, other countries' bond spreads decouple from the Greek bond spread. Secondly, although Greece is not propagating a lot of shocks to others in a linear way, it is the key player in terms of non-linear contagion. The break in Greece' bond spread pricing function comes earlier, which makes other countries more likely to switch to the crisis regime. All other sample countries, except Greece itself, are subject to considerable contagion effect (i.e., their transition probabilities to the crisis regime all significantly increase when their neighbors are in crisis in the past state). For those countries, multi-country contagion plays a more important role than common risk factors and even country-specific fundamentals in determining their transition probabilities to the crisis regime.

The rest of this paper is organized as follows: Section 2 introduces the mutual exciting regime-switching model. Section 3 discusses the Bayesian estimation procedure and inference. Section 4 presents the empirical application. Section 5 concludes.

1.2 Mutual Exciting Regime-Switching Model and Contagion

Consider the following regime-switching model:

$$\begin{aligned} y_{1t} &= \alpha_1(s_{1t}) + \boldsymbol{x}_{1t}'\boldsymbol{\beta}_1(s_{1t}) + h_{1t}^{1/2}(s_{1t})\varepsilon_{1t} \\ &\vdots \\ y_{nt} &= \alpha_n(s_{nt}) + \boldsymbol{x}_{nt}'\boldsymbol{\beta}_n(s_{nt}) + h_{nt}^{1/2}(s_{nt})\varepsilon_{nt} \end{aligned} \tag{1.1}$$

where y_{it} is a performance indicator for country i at time t for $i = 1,\ldots,n, t = 1,\ldots,T$. x_{it} is a $k \times 1$ vector of explanatory variables for country i at time t, which includes exogenous observed common factors and country specific variables. The country-specific state variable s_{it} is a hidden discrete stochastic process, which is unobserved. And the process $\{s_{it}\}$ is assumed to be irreducible and aperiodic first order Markov chain with finite state space $\{0,\ldots,K-1\}$. Different realizations of s_{it} admit different dynamics in the mean and in the variance. It is assumed that $\boldsymbol{\varepsilon_t} \sim \mathcal{N}(\mathbf{0},\Sigma)$, where $\boldsymbol{\varepsilon_t} = (\varepsilon_{1t},\ldots,\varepsilon_{nt})' \in \mathbb{R}^n$ for $t = 1,\ldots,T$ and Σ is a n-dimensional positive definite matrix with potentially non-zero off-diagonal entries. The variance of the random disturbance term $h_i(s_{it})\varepsilon_{it}$ depends on the realization of states via the volatility multiplier $h_{it}(s_{it})$. As for the parameters in the mean equations, the intercepts $\alpha_i(s_{it})$ and slopes $\boldsymbol{\beta}_i(s_{it})$ are also allowed to vary over the realizations of $\{s_{it}\}$.

Inter-dependence and contagion are distinguished in the model. While the inter-dependence is captured by the exposure to common factors, normal time interactions, and the non-zero off-diagonal elements in Σ, contagion is introduced by the mutual excitement component in the country-specific hidden stochastic process $\{s_{it}\}$. The evolution of $\{s_{it}\}$ is sufficiently described by the $K \times K$ time-varying transition matrix, which are governed by some exogenous variables $\boldsymbol{z}_{it}$ and all countries' past states $\boldsymbol{s}_{t-1} = (s_{1t-1},\ldots,s_{nt-1})$. We say there is contagion from j to i if the regime transition probability to crisis for i increases when j is in crisis in the last period, after controlling for other mechanisms that drive the switching process. For simplicity, in this paper, I only discuss the two-regime [1] ($K = 2$) case so that we can have a natural crisis regime and normal regime distinction. But generalization to more than two regimes is straightforward. I let only the first lagged states to enter the transition equation so that the vector of states $\boldsymbol{s}_{it} = (s_{1t},\ldots,s_{nt})$ still has the Markov property.[2] The transition

[1] Allowing for more regimes imposes no theoretical difficulties. But it could be computationally challenging when the number of countries n is big.

[2] More recent methodologies like forward-filtering backward sampling (FFBS) algorithm (Frühwirth-Schnatter (2006)) could be applied to allow for richer interaction pattern in different chains. This could

probability from state k to state l ($l,k = 0,1$ where $s = 0$ denotes the normal state and $s = 1$ denotes the crisis state) is specified as follows:

$$P(s_{it} = l \mid s_{it-1} = k, \boldsymbol{z}_{it}, \boldsymbol{s}_{-it-1}) = P^{i}_{kl,t}(\boldsymbol{z}_{it}, \boldsymbol{s}_{-it-1}), \quad \text{where } l,k = 0,1 \tag{1.2}$$

where $\boldsymbol{s}_{-it-1}$ is the vector of states for countries other than i at $t-1$. For the two-regime case, the unit-specific unobserved state variables follow a probit specification as in Equation 1.3. This can be generalized using a logit model if there are more than two regimes. Directed contagion effect from j to i is characterized by a positive λ_{ij}.

$$s_{it} = \begin{cases} 0 \text{ if } u_{it} < \boldsymbol{z}'_{it}\boldsymbol{\gamma_i} + \sum_{j \neq i} \lambda_{ij} s_{jt-1} \\ 1 \text{ if } u_{it} \geqslant \boldsymbol{z}'_{it}\boldsymbol{\gamma_i} + \sum_{j \neq i} \lambda_{ij} s_{jt-1} \text{ where } u_{it} \sim \mathcal{N}(0,1) \end{cases} \tag{1.3}$$

The model is different from the threshold model studied in Metiu (2012), where there is contagion is once a threshold is crossed. Here contagion is probabilistic rather than certain.

For the identification of a regime-switching model, one needs to deal with the label switching problem. A common way to achieve identification is to impose constraints on the parameters. This is used a lot in macroeconomic literature, and different regimes can have natural interpretations. In the empirical literature on contagion, the normal and the crisis regimes are often identified by different levels of asset returns' volatilities (Corsetti et al. (2005), Dungey* et al. (2005)). Given that line of reasoning, one reasonable identification restriction for the above contagion model is

$$h_{it}(s_{it} = 0) = 1 \text{ and } h_{it}(s_{it} = 1) > 1 \tag{1.4}$$

where $h_{it}(s_{it} = 1)$ is the volatility multiplier parameter.[3] This identification restriction does not impose an increase in the exposures to factors or a jump in the intercept. Whether a crisis state is associated with a significant break in the pricing function is left to be found out. Of course, this is not the only plausible identification restriction. Different restrictions could be applied, depending on the problem at hand.

be an interesting extension since letting the transition probability to depend on the whole path of the chain could be used to accommodate more interesting phenomena. For example, different duration of past bad states might change the transition probability by a different extent. Maybe some smoothing functions could be applied to summarise the information contained in the past state.

[3] For a two-regime case, another popular way to parameterize the regime-dependent volatility is to use $(1 + v_i * S_{it})\varepsilon_{it}$. And v_i can be interpreted as the proportional increase in volatility in the crisis state. Am equivalent identification restriction for such parameterization is $v_i > 0$.

1.3 Bayesian Inference by Gibbs Sampling

Putting everything together, for a two-regime case we have:

$$
\begin{aligned}
y_{it} &= \alpha_i(s_{it}) + \boldsymbol{x}'_{it}\boldsymbol{\beta}_i(s_{it}) + h_{it}^{1/2}(s_{it})\varepsilon_{it} \text{ for } i = 1,\ldots,n \\
s_{it} &= \begin{cases} 0 \text{ if } -u_{it} \geqslant \boldsymbol{z}'_{it}\boldsymbol{\gamma_i} + \sum_{j\neq i}\lambda_{ij}s_{jt-1} \\ 1 \text{ if } -u_{it} < \boldsymbol{z}'_{it}\boldsymbol{\gamma_i} + \sum_{j\neq i}\lambda_{ij}s_{jt-1} \text{ for } i = 1,\ldots,n \end{cases} \\
\boldsymbol{\varepsilon_t} &= \begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix} \sim \mathcal{N}(\mathbf{0},\Sigma) \text{ and } u_{it} \sim \mathcal{N}(0,1) \text{ for } i = 1,\ldots,n
\end{aligned}
\tag{1.5}
$$

The density function of observed performance variables conditional on states and all the parameters in the model, can be factorized as

$$
f(Y_T \mid S_T, X, \boldsymbol{\theta}) = f(Y_1 \mid S_1, X, \boldsymbol{\theta}) \prod_{t=2}^{T} f(\boldsymbol{y}_t \mid Y_{t-1}, S_t, X, \boldsymbol{\theta}) \tag{1.6}
$$

where $Y_t = (\boldsymbol{y}_1, \ldots \boldsymbol{y}_t)$ is the history of $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{nt})$ up to time t, $X = (\boldsymbol{x}'_{11}, \ldots, \boldsymbol{x}'_{1t}, \ldots, \boldsymbol{x}'_{n1}, \ldots, \boldsymbol{x}'_{nt})'$ is the matrix of exogenous regressors, $Z = (\boldsymbol{z}'_{11}, \ldots, \boldsymbol{z}'_{1t}, \ldots, \boldsymbol{z}'_{n1}, \ldots, \boldsymbol{z}'_{nt})'$ is the matrix of exogenous drivers of the regime switching process, $S_t = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_t)$ is the history of the states $\boldsymbol{s}_t = (s_{1t}, \ldots, s_{nt})$ up to time t, and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is the collection of parameters in the model. We collect all the parameters in the main equation in $\boldsymbol{\theta}_1 = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{h}, \Sigma)$ and all the parameters in the auxiliary regime-switching equation in $\boldsymbol{\theta}_2 = (\boldsymbol{\gamma}, \boldsymbol{\lambda})$.[4] In full, the joint density of the observations and states, is simply the product of the conditional density given states and the density of the states,

$$
f(Y_T, S_T \mid X, \boldsymbol{\theta}) = f(Y_T \mid S_T, X, \boldsymbol{\theta}_1) \prod_{t=2}^{T} P(\boldsymbol{s}_t \mid \boldsymbol{s}_{t-1}, Z, \boldsymbol{\theta}_2) \times P(\boldsymbol{s}_1) \tag{1.7}
$$

Direct calculations of the joint likelihood function are messy since brute force marginalization of Equation 1.7 involves 2^{n*T} summations over all possible state sequences $\{\boldsymbol{s}_t\}_{t=1}^{T}$. In this paper, Bayesian inference by Gibbs sampling as described in Albert and Chib (1993) and Kaufmann (2015) is applied to avoid the messy calculations involved in the direct evaluation of the joint likelihood function. For the simulation-based Bayesian procedure, unobserved

[4] $\boldsymbol{\alpha} = (\alpha_{1,0}, \alpha_{1,1}, \ldots, \alpha_{n,0}, \alpha_{n,1}), \boldsymbol{\beta} = (\boldsymbol{\beta}_{1,0}, \boldsymbol{\beta}_{1,1}, \ldots, \boldsymbol{\beta}_{n,0}, \boldsymbol{\beta}_{n,1})$. For a two-regime model, s_{it} is a dummy variable so that $\alpha_{i,1}$ and $\boldsymbol{\beta}_{i,1}$ correspond to the level shift and slope shift respectively.

states are treated as unknown parameters. And they can be simulated given other parameters in the model by Gibbs sampling.

We define the vector of latent indexes governing the transition process as $\boldsymbol{s}_t^* = (s_{1t}^*, \ldots, s_{nt}^*)$, where $s_{it}^* = \boldsymbol{z}_{it}'\boldsymbol{\gamma_i} + \sum_{j \neq i} \lambda_{ij} s_{jt-1} + u_{it}$. A key step in the procedure is to augment the data by $\boldsymbol{s}_t^*$ (i.e., the latent indexes $\boldsymbol{s}_t^*$ are also treated as unknown parameters). As a result of the data augmentation, the full parameters needed to be estimated are $\boldsymbol{\psi} = \{\boldsymbol{\theta}, S_T, S_T^*\}$, where $S_T = \{\boldsymbol{s}_t\}_{t=1}^T$ and $S_T^* = \{\boldsymbol{s}_t^*\}_{t=1}^T$ are the history of all states and the history of all latent indexes respectively. Our objective is to derive a Markov chain such that its limiting distribution is the joint distribution of interest. Let us divide the parameter set as $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \boldsymbol{\psi}_3, \boldsymbol{\psi}_4, \boldsymbol{\psi}_5, \boldsymbol{\psi}_6)$ where

$$\begin{aligned} \boldsymbol{\psi}_1 &= \{\boldsymbol{\alpha}, \boldsymbol{\beta}\} \\ \boldsymbol{\psi}_2 &= \{\Sigma\} \\ \boldsymbol{\psi}_3 &= \{\boldsymbol{h}\} \\ \boldsymbol{\psi}_4 &= \{S_T\} \\ \boldsymbol{\psi}_5 &= \{S_T^*\} \\ \boldsymbol{\psi}_6 &= \{\boldsymbol{\gamma}, \boldsymbol{\lambda}\} \end{aligned} \tag{1.8}$$

Let $[. \mid .]$ denotes the conditional distribution. The joint posterior distribution of $\boldsymbol{\psi}$ leads to very tractable conditional structure. And to sample from the posterior distribution, we iterate over the following steps:

1. Specifying arbitrary initial values $\boldsymbol{\psi}^0$ and set $i = 1$.
2. Cycle through the full conditionals by drawing
   - $\boldsymbol{\psi}_1^i$ from $[\boldsymbol{\psi}_1 \mid \boldsymbol{\psi}_2^{i-1}, \boldsymbol{\psi}_3^{i-1}, \boldsymbol{\psi}_4^{i-1}, \boldsymbol{\psi}_5^{i-1}, \boldsymbol{\psi}_6^{i-1}]$
   - $\boldsymbol{\psi}_2^i$ from $[\boldsymbol{\psi}_2 \mid \boldsymbol{\psi}_1^{i-1}, \boldsymbol{\psi}_3^{i-1}, \boldsymbol{\psi}_4^{i-1}, \boldsymbol{\psi}_5^{i-1}, \boldsymbol{\psi}_6^{i-1}]$
   - $\boldsymbol{\psi}_3^i$ from $[\boldsymbol{\psi}_3 \mid \boldsymbol{\psi}_1^{i-1}, \boldsymbol{\psi}_2^{i-1}, \boldsymbol{\psi}_4^{i-1}, \boldsymbol{\psi}_5^{i-1}, \boldsymbol{\psi}_6^{i-1}]$
   - $\boldsymbol{\psi}_4^i$ from $[\boldsymbol{\psi}_4 \mid \boldsymbol{\psi}_1^{i-1}, \boldsymbol{\psi}_2^{i-1}, \boldsymbol{\psi}_3^{i-1}, \boldsymbol{\psi}_5^{i-1}, \boldsymbol{\psi}_6^{i-1}]$
   - $\boldsymbol{\psi}_5^i$ from $[\boldsymbol{\psi}_5 \mid \boldsymbol{\psi}_1^{i-1}, \boldsymbol{\psi}_2^{i-1}, \boldsymbol{\psi}_3^{i-1}, \boldsymbol{\psi}_4^{i-1}, \boldsymbol{\psi}_6^{i-1}]$
   - $\boldsymbol{\psi}_6^i$ from $[\boldsymbol{\psi}_6 \mid \boldsymbol{\psi}_1^{i-1}, \boldsymbol{\psi}_2^{i-1}, \boldsymbol{\psi}_3^{i-1}, \boldsymbol{\psi}_4^{i-1}, \boldsymbol{\psi}_5^{i-1}]$

   where the conditioning on Y_T, X and Z are suppressed.
3. Let $i = i + 1$ and to back to the previous step.

The process generates a Markov chain, which under mild conditions (Tierney (1994)) has the joint distribution of interest as the limiting distribution. The first M draws have to be discarded, which is called "burn-in". After the "burn-in-period", the simulated values $(\boldsymbol{\psi}_1^i, \boldsymbol{\psi}_2^i, \boldsymbol{\psi}_3^i, \boldsymbol{\psi}_4^i, \boldsymbol{\psi}_5^i, \boldsymbol{\psi}_6^i)$ for $i = M+1, \ldots, M+K$ can be treated as approximated sample from from the joint posterior distribution. To initialize the sampler, one need initial values. I choose initial values with minimal prior information. The full conditionals and the choice of priors are provided in the appendix. Once we have the posteriors, we can then obtain the credible interval, which is analogous to the confidence interval in frequentist inference, for each parameter of interest.

1.4 Revisit Sovereign Credit Risk Contagion in the Eurozone

The econometric framework is applied to revisit the sovereign credit risk contagion in the Eurozone. We first discuss the data and its properties. We then examine the drivers of the regime-switching process. In particular, we are interested in the testing and quantification of multi-country contagion.

1.4.1 Data Description

Sovereign credit risk is measured by government bond yield spreads (relative to benchmark country Germany). Daily sovereign bond spreads of six Eurozone countries, including Greece, Ireland, Portugal, Spain, Italy (GIPSI countries), and France, are constructed using the difference between the 10-year sovereign bond yields of these countries and that of Germany. The daily data spans from 12/02/2008 to 01/12/2011 and are downloaded from Thomson Reuters Eikon. In the spirit of Caporin et al. (2018), we deliberately end the sample before the European Central Bank (ECB) announced the Long-Term Refinancing Operations (LTRO) to avoid the clustering of switchings due to the intervention.

Table 7 presents key macroeconomic fundamentals that affect credit conditions for the sample countries. Germany has the highest credit ratings, the best average fiscal position, and the highest GDP growth within the sample period. Sovereign bonds issued by the German government have very low yields and are considered extremely safe. That justifies why using German yield as the benchmark when constructing the spread is the convention in the literature (Bernoth et al. (2012), Metiu (2012), De Santis (2014), etc). France has the second best credit ratings, and its sovereign bond yield remains low during the whole sample period.

Spain and Italy follow and have the next worse credit ratings, with Italy having a much higher public debt level but a better fiscal position. Then we have three countries exhibiting high credit risk, Ireland, Portugal, and Greece. Ireland has the worst fiscal position in the sample, and Greece has the highest public debt level among sample countries.

Figure 1.1 and Figure 1.2 show the 10-year sovereign bond spreads of the six sample countries during the sample periods in level and in first difference, respectively. Sovereign bond spreads start dropping from spring 2009 as global uncertainty decreases and countries recovering from the global financial crisis. However, the spreads start to skyrocket at the end of 2009, when the Greek problem reveals. The rise is so sharp that it is hard to reconcile with the gradual deterioration of fundamentals, justifying the use of a regime-switching model to accommodate such breaks. Regime switch in the sovereign credit risk pricing equation for euro area peripheral countries is empirically supported (e.g., Favero and Missale (2012), Delatte et al. (2017)). I aim to go beyond them by allowing cross-sectional interaction in the regime-switching process and using the framework to test and quantify the non-linear contagion effects among sovereigns. Another important feature of Figure 1.1 is that the sovereign bond spreads show high persistency during the sample period. It is necessary to verify that these variables are stationary since the lack of stationarity will lead to deceptive results. Table 8 presents the results of stationarity tests for the sovereign bond yield spreads of all six sample countries. Augmented Dickey-Fuller test and Phillips-Perron test are both applied, showing these series are difference-stationary. Hence we use a first difference specification.

Based on the empirical literature focusing on the factors that determine individual sovereign credit spread (Edwards (1983), Edwards (1986), Duffie et al. (2003), Longstaff et al. (2011)), the factors affecting the sovereign bond yield spreads are associated with (1) common risk factors, (2) spillover effect, (3) country-specific risk factors, and (4) contagion risk. As for common risk factors, it is found that market risk appetite and uncertainty play an important role in the determination of sovereign risk (Baek et al. (2005)). I use two variables to proxy the market appetite and uncertainty in the euro area. The first one is the spread between the 3-month Euro Interbank Offered Rate (Euribor) and the Euro Overnight Index Average (EONIA). The second one is the VSTOXX Index, which is a forward-looking measure designed to reflect the market's expectations of future volatility in the euro area. For both variables, we use the first lag. To allow for the spillover effect, I use the first two lagged sovereign bond spreads from other countries in the sample. Country-specific default risk is determined by some low-frequency macroeconomic fundamental variables, including public debt/GDP ratio, fiscal balance/GDP ratio, GDP growth, and the current account/GDP

ratio. These low-frequency variables will drop out after taking the first difference of the daily data. As stated in the introduction section, contagion is defined a breakdown of the normal time transmission regime. In this model, the contagion risk is captured by a rise in the country-specific probability of being in the crisis regime when others are in crisis in the last period. As in Ait-Sahalia et al. (2014) and Aït-Sahalia et al. (2015), the contagion is probabilistic rather than certain.

A country's regime-switching process also reflects exposure to common risk factors and country-specific fundamentals. This paper contributes by explicitly allowing for the role of multi-country contagion in the regime-switching process so that crises can spread contagiously across countries in a probabilistic way. Again, we use the two common factors described above to control for inter-dependence in the switching process. It has been documented that government debt has non-linear effect on sovereign bond spreads (Bernoth et al. (2012), Delatte et al. (2017)). Due to that reason, I include the country-specific Debt-to-GDP ratio in the switching equation. Because governments might adjust their debt-to-GDP ratio endogenously in response to shocks to credit risk, I use the Debt-to-GDP ratio observed a quarter ahead so that it is predetermined with respect to the bond spread.

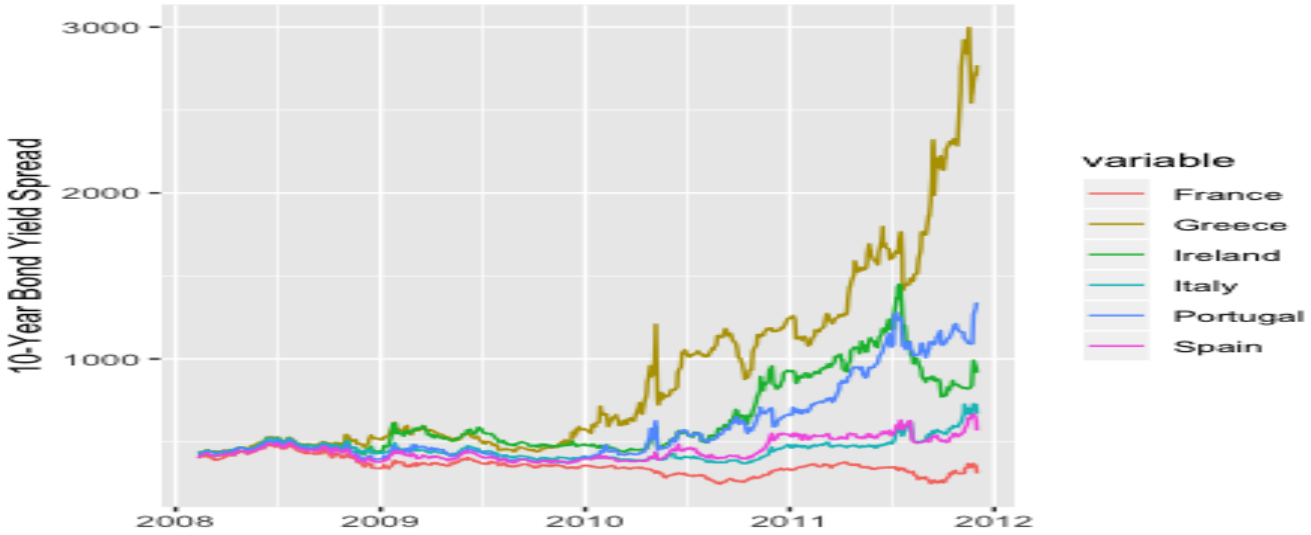


Figure 1.1 Daily 10-year sovereign bond spreads (in basis points)

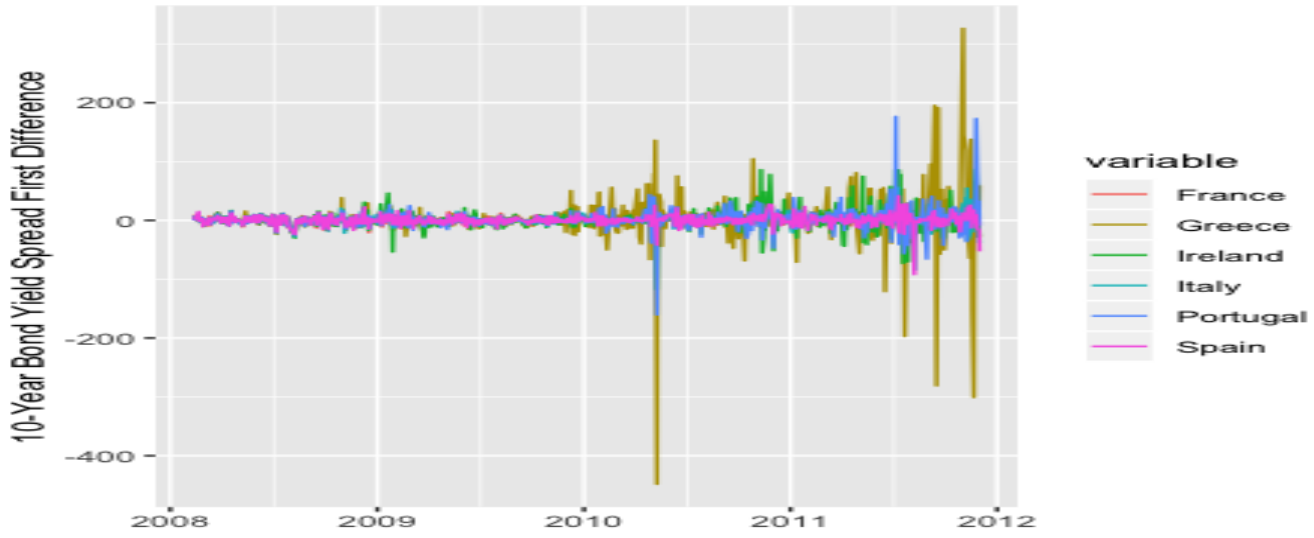


Figure 1.2 First-differenced daily 10-year sovereign bond spreads (in basis points)

1.4.2 Empirical Results

I implement the estimation methodology outlined above on the first-differenced 10-year sovereign bond yields. For the modeling of multi-country contagion effect, I adopt the formulation in Pesaran and Pick (2007) and aggregate the contagion effect from $N-1$ remaining countries.[5] To be more specific, the regime-switching equation (Equation 1.3) is modified to :

$$s_{it} = \begin{cases} 0 \text{ if } u_{it} < \boldsymbol{z}'_{it}\boldsymbol{\gamma_i} + \lambda_i I(\sum_{j\neq i}^{N} s_{jt-1}) \\ 1 \text{ if } u_{it} \geqslant \boldsymbol{z}'_{it}\boldsymbol{\gamma_i} + \lambda_i I(\sum_{j\neq i}^{N} s_{jt-1}) \text{ where } u_{it} \sim \mathcal{N}(0,1) \end{cases} \tag{1.9}$$

Under this formulation, the crisis indicator $I(\sum_{j\neq i}^{N} s_{jt-1})$ is a dummy variable that takes the value of one as long as any of the $N-1$ remaining countries are in a crisis state at $t-1$. In summary, the empirical specification is as follows:

$$\begin{aligned} &\Delta y_{it} = \alpha_i(s_{it}) + \boldsymbol{x}'_{it}\boldsymbol{\beta}_i(s_{it}) + h_{it}^{1/2}(s_{it})\varepsilon_{it} \text{ for } i = 1,\ldots,n \\ &s_{it} = \begin{cases} 0 \text{ if } u_{it} < \boldsymbol{z}'_{it}\boldsymbol{\gamma_i} + \lambda_i I(\sum_{j\neq i}^{N} s_{jt-1}) \\ 1 \text{ if } u_{it} \geqslant \boldsymbol{z}'_{it}\boldsymbol{\gamma_i} + \lambda_i I(\sum_{j\neq i}^{N} s_{jt-1}) \text{ where } u_{it} \sim \mathcal{N}(0,1) \end{cases} \\ &\boldsymbol{\varepsilon_t} \sim \mathcal{N}(\mathbf{0},\Sigma) \text{ and } u_{it} \sim \mathcal{N}(0,1) \text{ for } i = 1,\ldots,n \end{aligned} \tag{1.10}$$

[5] Estimating directed pairwise contagion as in Equation 1.3 poses no theoretical difficulties. The estimation results are available upon request.

where $\boldsymbol{x}_{it} = (\Delta f_{1t-1}, \Delta f_{2t-1}, \Delta y_{1t-1}, \Delta y_{1t-2}, \ldots, \Delta y_{Nt-1}, \Delta y_{Nt-2})'$ are the vector of explanatory variables described in section 4.1 in first difference. f_1 and f_2 correspond to the spread between Euribor and ENOIA, and the VSTOXX Index, respectively. $\Delta y_{it-1}, \Delta y_{it-2}$ for $i = 1, \ldots, N$ are the vector-autoregressive terms with two lags. The mean equation of Equation 1.10 is essentially a regime-dependent VARX model, and the structural model can be recovered by imposing identification assumptions on Σ. $\boldsymbol{z}_{it} = (1, \Delta f_{1t-1}, \Delta f_{2t-1}, Debt_{i,t_q-1}, S_{i,t-1})'$, together with $I(\sum_{j\neq i}^{N} s_{jt-1})$ gives the drivers of the regime-switching process. Notice that Debt-to-GDP ratio is observed at quarterly frequency, thus the variable has a different time subscript t_q. Prior distributions are provided in Table 9 and Table 10, which are used to initialize the Gibbs Sampler. I run 6000 iterations in total. The first 1000 "burn-in" iterations are discarded, and the 5000 iterations after that are treated as approximate sample from the joint posterior distribution.

Table 1.1 reports the posterior estimates of coefficients in the normal regime, and Table 1.2 reports the posterior estimates of the shift parameters (i.e., the changes in parameters when a country switches from the normal to crisis regime). The tables reveal that the sovereign bond spread pricing functions are highly regime-dependent as there there are striking shifts in market pricing behaviors. Our identification scheme is based on a rise in volatility in the crisis regime, and the economic magnitude of that is given by the volatility multiplier parameter h. Greek yield spread experience an almost eight-fold increase in volatility when it switches to the crisis regime. Other countries except for France also show high increases in their volatilities, with the posterior means of their volatility multiplier vary from 2 to 5. In the crisis regimes, all sample countries except France experience a significantly positive jump in the intercept. As the country with the best fundamental in our sample, France has a more "tranquil" crisis regime than others. The break of exposures to common risk factors is also worth noticing since the directions of the shifts are opposite to the sign of exposures in the normal regime. This might because the risk aversion and uncertainty both start falling since Spring 2009 while the euro area sovereign spreads begin to skyrocket after 2010. This suggests that the factor cannot help to interpret the sharp increases in euro area sovereign bond spreads during the European debt crisis. This phenomenon is also documented in De Santis (2014), where the author finds common risk factors stop being important determinants of European bond yields. The spillover pattern among sample countries also changes drastically from one regime to the other. The vector auto-regressive coefficients are much larger in magnitude in the crisis regime, indicating that regional risk spillovers explain more variations in the sovereign bond spreads during periods of crisis. This breakdown is associated with an

increase in interconnectedness that is beyond what can be explained by the normal time risk transmission mechanism.

	France(FR)	Spain(ES)	Italy(IT)	Portugal(PT)	Ireland(IE)	Greece(GR)
α	-0.05	0.12	**0.16**	**0.40**	**-0.17**	**0.42**
	[-0.18,0.13]	[-0.03,0.32]	[0.06,0.26]	[0.21,0.60]	[-0.44,-0.02]	[0.15,0.70]
β_1	**0.11**	**0.10**	**0.05**	0.07	**0.18**	**-0.28**
	[0.05,0.17]	[0.05,0.20]	[0.01,0.08]	[-0.01, 0.13]	[0.08,0.30]	[-0.36,-0.20]
β_2	**0.02**	**0.04**	**0.04**	**0.07**	0.01	-0.02
	[0.01,0.03]	[0.03, 0.05]	[0.03, 0.04]	[0.04, 0.08]	[-0.02,0.03]	[-0.03, 0.01]
$\beta_{FR,1}$	-0.01	**-0.07**	**-0.11**	**-0.26**	**-0.30**	-0.00
	[-0.03, 0.03]	[-0.09, -0.05]	[-0.13, -0.08]	[-0.28,-0.23]	[-0.35,-0.25]	[-0.06,0.05]
$\beta_{FR,2}$	0.02	0.03	0.03	**0.04**	**0.23**	**-0.10**
	[-0.02, 0.06]	[-0.02, 0.08]	[-0.01, 0.05]	[0.01, 0.08]	[0.18, 0.27]	[-0.15, -0.03]
$\beta_{ES,1}$	-0.01	**0.07**	0.02	**0.22**	-0.05	0.03
	[-0.04, 0.02]	[0.02,0.15]	[-0.04, 0.07]	[0.19,0.24]	[-0.08,0.00]	[-0.04, 0.08]
$\beta_{ES,2}$	-0.01	-0.00	**0.03**	0.02	**0.18**	**0.10**
	[-0.02, 0.01]	[-0.02,0.02]	[0.01,0.03]	[-0.01,0.04]	[0.15, 0.21]	[0.07, 0.13]
$\beta_{IT,1}$	**0.02**	-0.01	-0.01	**0.05**	**-0.02**	**-0.09**
	[0.01, 0.02]	[-0.03,0.00]	[-0.03,0.01]	[0.03, 0.07]	[-0.04, 0.01]	[-0.12,-0.06]
$\beta_{IT,2}$	**0.01**	0.00	**0.02**	**-0.02**	**0.01**	**0.08**
	[0.00, 0.01]	[-0.00,0.01]	[0.01, 0.02]	[-0.02,-0.01]	[0.00,0.01]	[0.06,0.10]
$\beta_{PT,1}$	**-0.21**	**-0.12**	-0.00	**-0.03**	**-0.16**	**0.06**
	[-0.24, -0.16]	[-0.14, -0.06]	[-0.03, 0.01]	[-0.06, -0.01]	[-0.20, -0.12]	[0.03, 0.12]
$\beta_{PT,2}$	**0.06**	-0.04	**-0.03**	**-0.15**	0.03	**0.09**
	[0.01, 0.08]	[-0.06, 0.00]	[-0.07, -0.01]	[-0.20, -0.12]	[-0.05, 0.06]	[0.06, 0.13]
$\beta_{IE,1}$	**0.13**	**0.18**	**0.07**	**0.14**	**0.15**	**-0.07**
	[0.11, 0.17]	[0.09, 0.23]	[0.05, 0.10]	[0.11, 0.17]	[0.10, 0.17]	[-0.10, -0.03]
$\beta_{IE,2}$	**0.04**	0.03	**0.03**	**0.11**	**-0.06**	**0.06**
	[0.02, 0.05]	[-0.00, 0.05]	[0.01, 0.05]	[0.09, 0.13]	[-0.09,-0.03]	[0.03, 0.08]
$\beta_{GR,1}$	**-0.04**	**-0.06**	**-0.03**	**-0.05**	**0.12**	0.01
	[-0.06, -0.03]	[-0.07, -0.05]	[-0.05, -0.02]	[-0.07, -0.00]	[0.09, 0.14]	[-0.01, 0.03]
$\beta_{GR,2}$	**-0.03**	-0.01	**-0.02**	**-0.02**	**-0.02**	**0.01**
	[-0.03, -0.02]	[-0.02, 0.01]	[-0.02, -0.01]	[-0.03, -0.02]	[-0.02, -0.01]	[0.00,0.01]

Table 1.1 95% credible intervals and posterior mean for the normal regime parameters. Note: β_1 is the coefficient on the lagged first-differenced spread between Euribor and ENOIA, β_2 is the coefficient on the lagged first-differenced VSTOXX Index. $\beta_{i,t-k}$ is the coefficient on Δy_{it-k}, for $i = FR, ES, IT, PT, IE, GR$ and $k = 1, 2$. Coefficients significant at the 95 % level are in bold.

	France(FR)	Spain(ES)	Italy(IT)	Portugal(PT)	Ireland(IE)	Greece(GR)
$\Delta\alpha$	0.27	**2.35**	**2.34**	**4.71**	**3.94**	**8.19**
	[-0.12,0.60]	[1.29,3.24]	[1.73,3.39]	[2.84,6.77]	[2.39,5.18]	[2.38,12.77]
$\Delta\beta_1$	-0.06	**-2.66**	**-1.34**	**-2.24**	**-2.93**	**-7.76**
	[-0.15,0.12]	[-2.87,-2.37]	[-1.77, -1.06]	[-3.03, -1.20]	[-3.67,-2.42]	[-9.31, -6.26]
$\Delta\beta_2$	-0.03	0.06	**-0.17**	0.12	-0.08	-0.10
	[-0.05, -0.00]	[-0.01, 0.22]	[-0.36, -0.05]	[-0.08,0.31]	[-0.25, 0.06]	[-0.41, 0.14]
$\Delta\beta_{FR,1}$	**0.33**	0.07	-0.07	**0.48**	**0.70**	**-3.29**
	[0.26, 0.41]	[-0.03, 0.19]	[-0.23, 0.19]	[0.23, 0.79]	[0.39, 0.91]	[-4.03, -2.63]
$\Delta\beta_{FR,2}$	0.01	**0.83**	**0.46**	**1.39**	**1.04**	**1.11**
	[-0.07, 0.11]	[0.59, 0.94]	[0.28,0.59]	[1.06, 1.62]	[0.47,1.33]	[0.33, 2.13]
$\Delta\beta_{ES,1}$	**0.21**	**-0.52**	**-0.12**	**-1.13**	**-1.33**	**3.07**
	[0.15, 0.25]	[-0.63, -0.37]	[-0.25, -0.01]	[-1.34, -0.79]	[-1.63, -0.65]	[1.75, 3.74]
$\Delta\beta_{ES,2}$	**-0.13**	**-0.17**	**-0.31**	-0.02	**-0.20**	**-0.26**
	[-0.15, -0.11]	[-0.26, -0.07]	[-0.42, -0.20]	[-0.16,0.05]	[-0.30, -0.10]	[-0.47, -0.03]
$\Delta\beta_{IT,1}$	**-0.07**	**0.21**	**0.20**	**0.20**	**0.20**	0.03
	[-0.10, -0.01]	[0.10, 0.32]	[0.12, 0.26]	[0.13, 0.27]	[0.10,0.34]	[-0.17,0.20]
$\Delta\beta_{IT,2}$	**0.03**	-0.03	**0.03**	0.04	0.03	-0.04
	[0.02, 0.05]	[-0.06, -0.00]	[0.01, 0.05]	[0.00, 0.10]	[-0.01,0.05]	[-0.10, 0.03]
$\Delta\beta_{PT,1}$	**0.71**	0.13	**-0.54**	0.23	**-0.42**	**-2.78**
	[0.62, 0.76]	[-0.05, 0.29]	[-0.79, -0.10]	[-0.16, 0.89]	[-0.70,-0.20]	[-3.17, -2.21]
$\Delta\beta_{PT,2}$	**0.13**	0.00	-0.15	**-1.39**	**-1.23**	**-2.17**
	[0.061, 0.23]	[-0.08,0.12]	[-0.47, 0.25]	[-1.64, -1.07]	[-1.47,-0.90]	[-2.68, -1.60]
$\Delta\beta_{IE,1}$	**-0.29**	**-0.31**	-0.20	**0.66**	**0.69**	**1.86**
	[-0.36, -0.25]	[-0.43, -0.23]	[-0.56, 0.00]	[0.31, 0.89]	[0.30, 1.03]	[1.46, 2.23]
$\Delta\beta_{IE,2}$	**-0.12**	0.02	0.06	**-0.21**	0.02	**-0.31**
	[-0.14, -0.11]	[-0.03, 0.08]	[-0.00, 0.12]	[-0.27,-0.12]	[-0.08, 0.10]	[-0.60, -0.07]
$\Delta\beta_{GR,1}$	**0.16**	**0.18**	-0.00	**0.22**	**0.22**	0.18
	[0.13, 0.19]	[0.12,0.23]	[-0.09, 0.08]	[0.16, 0.27]	[0.10,0.32]	[-0.12, 0.52]
$\Delta\beta_{GR,2}$	**0.03**	**-0.05**	**0.03**	0.06	**-0.07**	**-0.14**
	[0.01, 0.05]	[-0.08, -0.01]	[0.02, 0.05]	[0.00, 0.12]	[-0.11, -0.04]	[-0.19, -0.06]
h	**1.11**	**2.71**	**3.65**	**4.89**	**4.46**	**7.80**
	[1.00,1.36]	[2.32,3.21]	[3.10,4.30]	[4.31, 5.54]	[3.93, 5.07]	[6.90, 8.82]

Table 1.2 95% credible intervals and posterior mean for the shift parameters in the mean equation.
Note: A typical shift parameter $\Delta\beta = \beta(s=1) - \beta(s=0)$, is the change of that parameter from normal regime to crisis regime. Coefficients significant at the 95 % level are in bold.

To look closer into the roles played by countries in the regional risk spillover, I calculate the variance decomposition network as in Diebold and Yılmaz (2014) for the two regimes. For the orthogonalization of shocks, instead of using Cholesky decomposition, which requires an ordering of variables, I adopt the generalized impulse response function from Pesaran and Shin (1998). Table 1.3 and Table 1.4 report the variance decomposition network when all countries are in normal and crisis regime, respectively. The tables summarize all dependencies up to lag h by means of the forecast error variance decomposition. A typical element on the ith row and jth column gives the percentage of h-step ahead forecast error variance of Δy_{it} that is due to innovations in Δy_{jt}. The row sums are ones as a result of normalization,

and the kth column sum gives the from-degree of country k (i.e., total spillover from k). Comparing the variance decomposition networks during normal and crisis regime reveals some interesting features. Spain, Italy, and Portugal are the most important systemically risk contributors. On the other hand, the two countries that are considered as the origin of the "fever", do not propagate a lot of risk to the system. This result is also documented in Caporin et al. (2018) and Dumitru and Holden (2019). [6] On top of their findings, we find that while Italy is the hub of the network in the normal regime, and its role gets replaced by Portugal in the crisis regime. This might because Portugal is the systemically most important debtor based on the network structure of debt among sample countries. When the system is in stress, the credit risk from Portugal quickly spreads to other countries via debt exposure. On the other hand, Greece plays a less important role in directly propagating shocks to others. This might be explained by the fact that investors start to isolate Greek bonds from other countries when the Greek default is inevitable. However, it will be seen later that Greek plays an important role in non-linear contagion as its break comes earlier and makes other countries more likely to switch to the crisis regime. Overall, we confirm that there is strong evidence of parameters instability during our sample period. Regional spillovers gain importance during periods of stress, with Italy, Spain, and Portugal playing important roles in directly spill over risk to others. Common risk factors, on the other hand, fail to explain the sharp increases in euro area sovereign bond spreads.

	France(FR)	Spain(ES)	Italy(IT)	Portugal(PT)	Ireland(IE)	Greece(GR)
France(FR)	14.82	20.56	28.54	15.90	12.81	7.38
Spain(ES)	12.52	21.20	28.75	16.55	13.61	7.36
Italy(IT)	12.38	20.50	29.07	16.56	13.57	7.92
Portugal(PT)	11.48	19.91	27.91	18.25	14.53	7.93
Ireland(IE)	11.43	19.76	27.65	17.42	15.85	7.88
Greece(GR)	10.98	18.40	27.57	16.52	13.60	12.92
From-Degree	73.61	120.33	169.50	101.19	83.97	51.40

Table 1.3 Variance decomposition network when all countries are in normal regime. The prediction horizon is 5 days.

[6] Caporin et al. (2018) documents Italy's role as the hub of the network of sovereign contagion during the European debt crisis. Spain and Portugal's important roles are also found in Dumitru and Holden (2019). Both studies find other countries' bond spreads decouple from the Greek bond spread.

	France(FR)	Spain(ES)	Italy(IT)	Portugal(PT)	Ireland(IE)	Greece(GR)
France(FR)	17.17	15.45	18.94	32.27	11.57	4.60
Spain(ES)	8.24	34.22	25.23	18.41	11.72	2.19
Italy(IT)	9.20	15.50	22.33	35.23	13.07	4.66
Portugal(PT)	15.00	5.46	12.07	46.24	17.38	3.85
Ireland(IE)	11.41	17.66	16.86	34.19	16.05	3.83
Greece(GR)	8.22	13.18	21.57	29.26	22.36	5.42
From-Degree	69.24	101.47	117.0	195.59	92.15	24.55

Table 1.4 Variance decomposition network when all countries are in crisis regime. The prediction horizon is 5 days.

Table 1.5 reports the posterior estimates of coefficients in the auxiliary regime-switching equation. The first important observation is that common risk factors do not play a role in determining the probabilities of regime-switching. For a country, its lagged Debt-to-GDP ratio, past state, and other countries' past states (the contagion component) determine its transition probability. As documented in Bernoth et al. (2012) and Delatte et al. (2017), government debt level has an non-linear effect on sovereign bond spreads. For all sample countries except Italy, a higher Debt-to-GDP ratio corresponds to a higher probability of entering the crisis regime. A country's own past state largely affects the transition probability. A normal state is much more likely to be followed by a normal state, while a crisis state increases the probability of getting another crisis state in the next period by a large margin. Apart from a country's own past states, other countries' past states also significantly affect the country's transition probability. There is strong evidence of multi-country contagion. Except for Greece, other countries' transition probabilities to crisis regime all significantly increase when at least one of their neighbors are in crisis in the past state. Greece is not affected by multi-country contagion since its break comes earlier than others, and conditional on its own past states, other countries' past states do not have an additional effect on its transition probability. For France and Spain, the contagion effect on transition probability is equivalent to an increase in the Debt-to-GDP ratio of around 20%. Portugal and Ireland are subject to an even large contagion effect, and its effect on transition probability is equivalent to an increase in the Debt-to-GDP ratio of around 40%. To better interpret the intensity of contagion, I conduct a partial effect analysis since the coefficients alone in the non-linear regime-switching process could be less indicative. Table 1.6 shows the partial effect (PE) of contagion on transition probabilities when countries' other switching variables are at different percentiles. Since a country's own past state is discrete, we separately analyze the partial effect on transition probability from normal to crisis and crisis to crisis. The economic

magnitude of the contagion effect is large. It contributes to more than a 10% increase in transition probability from normal to crisis regime for France, Spain, Portugal, and Ireland, given their Debt-to-GDP ratio at any percentiles considered. Especially for Portugal and Ireland, multi-country contagion is associated with more than 25% increases in transition probabilities when their Debt-to-GDP ratio is at the median level. The incremental effect of contagion is smaller when Portugal's debt level is high, as its own fundamentals now contribute a lot to the switching. The contagion effect on transition probability from crisis to crisis regime is also sizeable, although smaller for most countries. The smaller partial effect is because the probability of staying in the crisis regime is already high, making the incremental effect of contagion smaller in magnitude. Overall, I find strong evidence of multi-country contagion during the sample period. It plays a more important role than common risk factors and even country-specific fundamentals in determining the transition probability to crisis regime.

	France(FR)	Spain(ES)	Italy(IT)	Portugal(PT)	Ireland(IE)	Greece(GR)
γ_1	**-1.895**	**-2.071**	**-2.161**	**-1.672**	**-1.841**	**-1.648**
	[-1.939,-1.855]	[-2.094, -2.047]	[-2.214, -2.076]	[-1.738, -1.614]	[-1.897, -1.810]	[-1.717,-1.621]
γ_2	**-0.008**	0.007	0.005	-0.006	-0.000	0.008
	[-0.0168, -0.002]	[-0.004, 0.016]	[-0.002, 0.012]	[-0.015,0.005]	[-0.011, 0.014]	[-0.002, 0.014]
γ_3	0.001	-0.000	0.002	-0.001	-0.001	-0.002
	[-0.001,0.004]	[-0.002, 0.001]	[-0.001, 0.004]	[-0.002, 0.001]	[-0.003, 0.001]	[-0.003,-0.000]
γ_4	**0.028**	**0.026**	-0.004	**0.027**	**0.019**	**0.025**
	[0.025, 0.030]	[0.023, 0.031]	[-0.007, 0.001]	[0.026, 0.028]	[0.018, 0.020]	[0.021, 0.027]
γ_5	**1.586**	**1.896**	**2.407**	**0.827**	**1.353**	**2.052**
	[1.458,1.741]	[1.732,2.133]	[2.312, 2.484]	[0.732, 0.985]	[1.254,1.489]	[1.964, 2.128]
λ	**0.444**	**0.410**	**0.531**	**0.872**	**0.671**	0.041
	[0.337, 0.509]	[0.339,0.455]	[0.469, 0.590]	[0.794,0.922]	[0.582, 0.748]	[-0.046, 0.129]

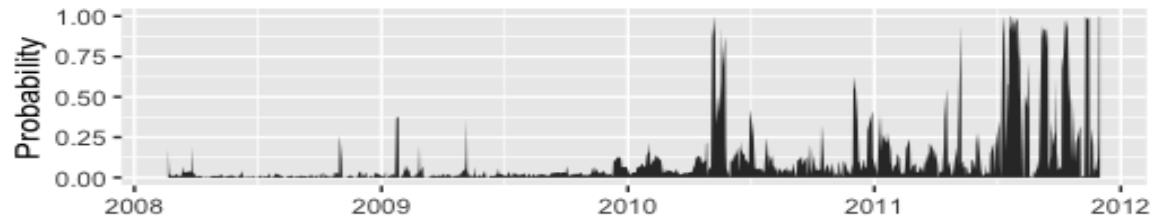
Table 1.5 95% credible intervals and posterior mean of the coefficients in the auxiliary regime-switching equation.

Note: γ_1 is the constant of the switching equation, γ_2 is the coefficient on the lagged first-differenced spread between Euribor and ENOIA, γ_3 is the coefficient on the lagged first-differenced VSTOXX Index, γ_4 is the coefficient on the lagged Debt-to-GDP ratio and γ_5 is the coefficient on own past state. λ is the coefficient of the contagion effect. Coefficients significant at the 95 % level are in bold.

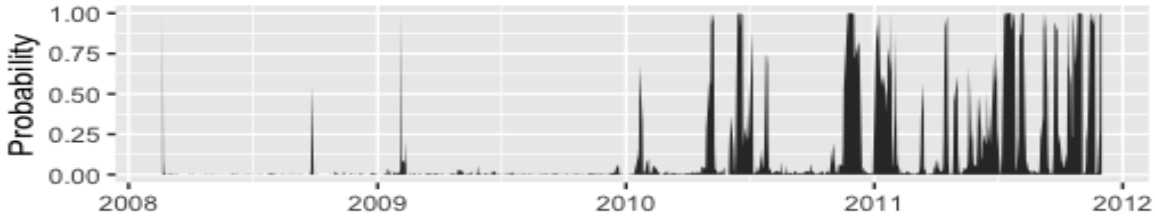
	Coefficient (λ_i)	PE at 50^{th}	PE at 75^{th}	PE at 90^{th}
(1) Normal to Crisis				
France(FR)	0.444	14.50%	13.60%	13.41%
Spain(ES)	0.410	14.49%	15.59%	16.10%
Italy(IT)	0.531	3.62%	3.62%	3.62%
Portugal(PT)	0.872	20.76%	14.24%	8.57%
Ireland(IE)	0.671	25.02%	26.00%	25.19%
(1) Crisis to Crisis				
France(FR)	0.444	1.59%	1.28%	1.23%
Spain(ES)	0.410	6.05%	4.54%	3.53%
Italy(IT)	0.531	18.42%	18.42%	18.42%
Portugal(PT)	0.872	6.83%	3.60%	1.64%
Ireland(IE)	0.671	15.73%	8.85%	7.07%

Table 1.6 Partial effect (PE) of contagion at different percentiles of other regime-driving variables.
Note: For each sample country, the effect on transition probability from normal to crisis and crisis to crisis are given the the upper panel (1) and lower panel (2), respectively. I consider the lagged Debt-to-GDP ratio at their 50^{th}, 75^{th} and 90^{th} percentiles.

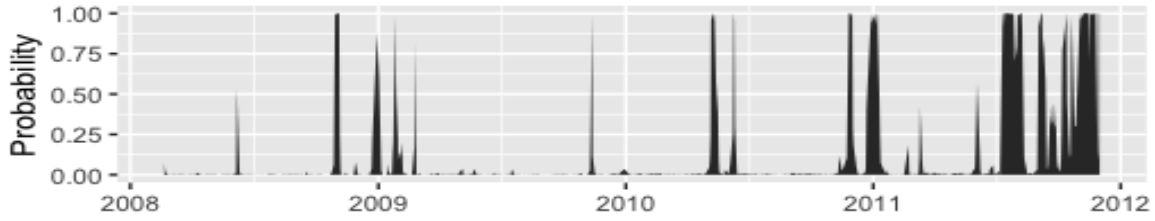
The model also produces country-specific probabilities for each regime. Figure 1.3 reports the crisis probability for each country. These figures provide some interesting results. Firstly, sample countries' regimes are not fully synchronized, and there is a considerable degree of heterogeneity in the cross-sectional regime-switching patterns. Secondly, the crisis regime is not continuous, hindering the usefulness of sample splitting type of contagion test as they rely on *ex-post* identification of the crisis regime, which implicitly assumes that the crisis regime is continuous and homogeneous. These figures also reveal why Greece is not subject to the contagion effect from others. At the end of 2009, while other countries are in the normal regime, Greece is the first country that enters the crisis regime as its trouble reveals in December 2009, when it admits its debts have reached 300*bn* euros, the highest in modern history. Thus, conditional on its own past states, other countries' states do not affect Greece's transition probability anymore.



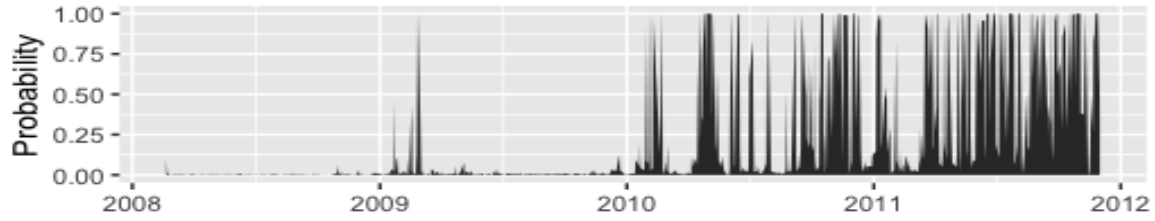
(a) France



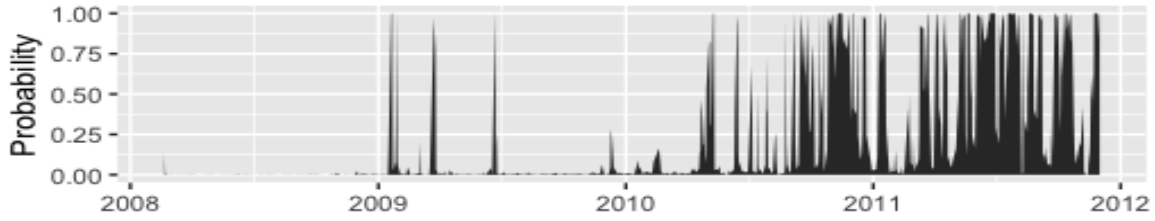
(b) Spain



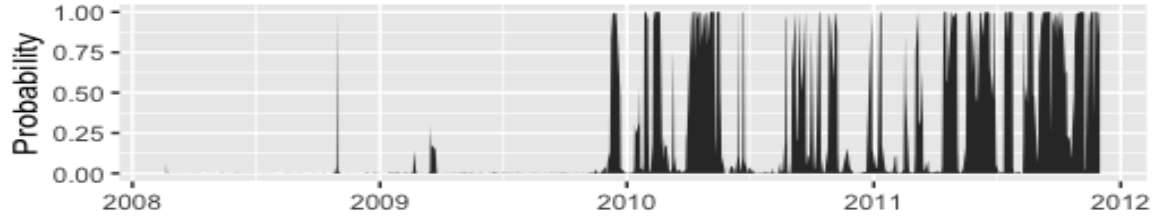
(c) Italy



(d) Portugal



(e) Ireland



(f) Greece

Figure 1.3 Country-specific smoothed probabilities for crisis regime.

1.5 Conclusion

This paper contributes by proposing a new methodological framework of the multi-country contagion problem. It develops a procedure to test and quantify contagion based on a mutual exciting regime-switching model. Contagion is defined as a rise in the transition probability to the crisis regime when other countries are in crisis in the past state. The model has several advantages. Firstly, it does not rely on *ex-post* identification of the crisis regime—this type of identification scheme implicitly assume that the crisis regime is continuous and homogeneous. However, from the empirical results, we can see that the country-specific crisis regime is neither continuous nor homogeneous. Secondly, different from the strand of literature that assumes the crisis regime is associated with some extreme values of the observed dependent variable, we let the crisis state to be an unobserved stochastic process. This is motivated by the fact that crises, in many cases, are more complicated and can not be sufficiently captured by the tail events of one particular value. Thirdly, we explicitly model multi-country contagion as a source of regime-switching. We can quantitatively analyze the roles of different mechanisms, especially multi-country contagion, in determining the transition process. Lastly, this framework accommodates a rich contagion pattern. In addition to a break in the mean equation, which is the parameters instability in the asset pricing equation, the model also accommodates a break in the second moment. The dynamics in the variance could be especially important for crisis episodes where increases in volatility are the major symptomatic.

The project also contributes from an empirical point of view. There is an extensive body of research examining sovereign bond prices in the context of the European debt crisis and whether there is a contagion effect remains the center of the debate. Empirical evidence is very much mixed. The empirical study in this paper provides some new important findings. First of all, there are striking shifts in market pricing behaviors. There is not only a jump in the intercept but also breaks in the exposures to common risk factors and the intensities of the regional spillover effect. It is vitally important to take into account this regime-dependent pricing behavior. Secondly, countries are subject to a strong contagion effect. Actually, contagion plays a more important role than common risk factors and country-specific fundamentals in determining their transition probabilities to the crisis regime.

Appendices

(A) Supplementary figures and tables

Country	Public Debt (%GDP, average 2008-2012)	GDP Growth(%, average 2008-2012)	Fiscal Position(%GDP, average 2008-2012)	Credit Ratings, 2012 (Moody's, Fitch, S&P)
Germany	70.2	0.7	-1.7	Aaa,AAA,AAA
France	75.9	0.6	-5.4	Aaa,AAA,AA+
Italy	108.7	0.6	-3.7	A3,A+,BBB+
Spain	47.5	0.6	-7.9	A3,AA-,A
Ireland	55.0	-0.3	-14.1	Ba1,BBB+,BBB+
Portugal	81.1	0.5	-7.4	Ba3,BBB-,BB
Greece	123.1	0.2	-10.9	Ca,CCC,CC

Table 7 Macroeconomic summary statistics for sample countries (source: ECB)

Variable	Augmented Dickey-Fuller		Phillips-Perron	
	Level	First Difference	Level	First Difference
FR	-2.16	-8.58***	-11.7	-753***
ES	-2.37	-9.41***	-12.5	-616***
IT	-0.09	-9.75***	-2.97	-682 ***
PT	-0.59	-8.98***	-2.44	-600***
IE	-1.83	-9.25***	-8.01	-730***
GR	0.58	-10.22***	-0.31	-611***

Table 8 Stationarity tests of 10-year sovereign bond spreads for six sample countries. ***, **, * denote the rejection of unit root hypothesis at the 1%, 5%, 10% level of significance, respectively.

(B) Prior Distributions

Parameters ($s_{it} = 0$)	Mean	Std.dev	Shift Parameters	Mean	Std.dev
α	0	10	$\Delta\alpha$	0.1	10
β_1	0	10	$\Delta\beta_1$	0	10
β_2	0	10	$\Delta\beta_2$	0	10
$\beta_{FR,1}$	0	10	$\Delta\beta_{FR,1}$	0	10
$\beta_{FR,2}$	0	10	$\Delta\beta_{FR,2}$	0	10
$\beta_{ES,1}$	0	10	$\Delta\beta_{ES,1}$	0	10
$\beta_{ES,2}$	0	10	$\Delta\beta_{ES,2}$	0	10
$\beta_{IT,1}$	0	10	$\Delta\beta_{IT,1}$	0	10
$\beta_{IT,2}$	0	10	$\Delta\beta_{IT,2}$	0	10
$\beta_{PT,1}$	0	10	$\Delta\beta_{PT,1}$	0	10
$\beta_{PT,2}$	0	10	$\Delta\beta_{PT,2}$	0	10
$\beta_{IE,1}$	0	10	$\Delta\beta_{IE,1}$	0	10
$\beta_{IE,2}$	0	10	$\Delta\beta_{IE,2}$	0	10
$\beta_{GR,1}$	0	10	$\Delta\beta_{GR,1}$	0	10
$\beta_{GR,2}$	0	10	$\Delta\beta_{GR,2}$	0	10
–			h	1.2	10

Table 9 Mean and standard deviation of priors on the main equation. The table shows the prior distributions for a typical country. We use the same sets of prior distributions for each sample country.

Parameters ($s_{it} = 0$)	Mean	Std.dev
γ_1	2	10
γ_2	0	10
γ_3	0	10
γ_4	0	10
γ_5	-2	10
λ	0	10

Table 10 Mean and standard deviation of priors on the auxiliary regime-switching equation. The table shows the prior distributions for a typical country. We use the same sets of prior distributions for each sample country.

(C) Sampling from Full Conditionals

To sample from the joint posterior distribution of full parameters $\boldsymbol{\psi}$ given data, we sample from the following conditional posteriors iteratively:

(1) $\boldsymbol{\psi}_1 = \{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$

We first rearrange Equation 1.1 as a linear regression model given other parameters,

$$\begin{aligned} & h_1^{-1}(s_{1t})(y_{1t} - \alpha_{1,0} - \alpha_{1,1}s_{1t} - \boldsymbol{x}_{1t}'\boldsymbol{\beta}_{1,0} - (\boldsymbol{x}_{1t} * s_{1t})'\boldsymbol{\beta}_{1,1}) = \varepsilon_{1t} \\ & \vdots \\ & h_n^{-1}(s_{nt})(y_{nt} - \alpha_{n,0} - \alpha_{n,1}s_{nt} - \boldsymbol{x}_{nt}'\boldsymbol{\beta}_{n,0} - (\boldsymbol{x}_{nt} * s_{nt})'\boldsymbol{\beta}_{n,1}) = \varepsilon_{nt} \end{aligned} \tag{11}$$

Let $\tilde{y}_{it} = h_i^{-1}(s_{it}) * y_{it}$, $\tilde{\boldsymbol{x}}_{it} = (h_i^{-1}(s_{it}), h_i^{-1}(s_{it}) * s_{it}, h_i^{-1}(s_{it})\boldsymbol{x}_{it}', h_i^{-1}(s_{it}) * s_{it} * \boldsymbol{x}_{it}')'$. Let $\tilde{Y}_i = (\tilde{y}_{i1}, \ldots, \tilde{y}_{iT})'$ and $\tilde{X}_i = (\tilde{\boldsymbol{x}}_{i1}, \ldots, \tilde{\boldsymbol{x}}_{iT})'$. Denote $\tilde{\boldsymbol{\beta}}_i = (\alpha_{i,0}, \alpha_{i,1}, \boldsymbol{\beta}_{i,0}', \boldsymbol{\beta}_{i,1}')'$ so that Equation 11 can be rewritten as:

$$\begin{aligned} \tilde{y}_{1t} &= \tilde{\boldsymbol{x}}_{1t}'\tilde{\boldsymbol{\beta}}_1 + \varepsilon_{1t} \\ &\vdots \\ \tilde{y}_{nt} &= \tilde{\boldsymbol{x}}_{nt}'\tilde{\boldsymbol{\beta}}_n + \varepsilon_{nt} \end{aligned} \tag{12}$$

Given ψ_2, ψ_3, ψ_4, with $\tilde{y}_{it}$ and $\tilde{\boldsymbol{x}}_{it}$ being observed, Equation 12 is a system of linear regressions with known variance covariance matrix. I use normal prior on $\tilde{\boldsymbol{\beta}}_i$

$$\tilde{\boldsymbol{\beta}}_i \sim \mathcal{N}(\tilde{\boldsymbol{\beta}}_i^0, \sigma_i^2 P_i^0) \text{ for } i = 1, \ldots, n \tag{13}$$

Subscript 0 and 1 represent the parameters for regime 0 and 1, respectively. Superscript 0 and 1 indicate the prior and posterior, respectively. σ_i^2 is the ith diagonal entry of Σ. The posterior distribution of $\tilde{\boldsymbol{\beta}}_i$ is given by:

$$\tilde{\boldsymbol{\beta}}_i \sim \mathcal{N}(\tilde{\boldsymbol{\beta}}_i^1, \sigma_i^2 P_i^1) \text{ for } i = 1, \ldots, n \tag{14}$$

where $P_i^1 = ((P_i^0)^{-1} + \tilde{X}_i'\tilde{X}_i)^{-1}$ and $\tilde{\boldsymbol{\beta}}_i^1 = P_i^1((P_i^0)^{-1}\tilde{\boldsymbol{\beta}}_i^0 + \tilde{X}_i'\tilde{Y}_i)$

(2) $\boldsymbol{\psi}_2 = \{\Sigma\}$

Given ψ_1, ψ_3, ψ_4, $\boldsymbol{y_t} \in \mathbb{R}^n$ for $t = 1, \ldots, T$ follows a multivariate normal distribution with known mean $\boldsymbol{\mu} \in \mathbb{R}^n$.

$$\boldsymbol{y_t} = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{nt} \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \tag{15}$$

The natural conjugate prior for a covariance matrix is the inverse Wishart (IW) prior, thus I impose:

$$\Sigma \sim \mathcal{IW}(\nu^0, \Lambda^0) \tag{16}$$

And the posterior of Σ is

$$\Sigma \sim \mathscr{IW}(T+\nu^0, S_\mu + \Lambda^0) \tag{17}$$

where $S_\mu = \sum_{t=1}^{T}(\boldsymbol{y_t} - \boldsymbol{\mu})(\boldsymbol{y_t} - \boldsymbol{\mu})'$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)' \in \mathbb{R}^n$ with $\mu_i = h_i^{-1}(s_{it})(\alpha_{i,0} - \alpha_{i,1}s_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_{i,0} - (\boldsymbol{x}_{it} * s_{it})'\boldsymbol{\beta}_{i,1})$.

(3) $\boldsymbol{\psi}_3 = \{\boldsymbol{h}\}$

$$y_{it} - \alpha_{i,0} - \alpha_{i,1}s_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_{i,0} - (\boldsymbol{x}_{it} * s_{it})'\boldsymbol{\beta}_{i,1} = h_i(s_{it})\varepsilon_{it} \tag{18}$$

Denote $\check{x}_i = y_{it} - \alpha_{i,0} - \alpha_{i,1}s_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta}_{i,0} - (\boldsymbol{x}_{it} * s_{it})'\boldsymbol{\beta}_{i,1}$. Given ψ_1, ψ_3 and ψ_4, and with $\check{x}_i$ being is observed, and a natural way to interpret equation h_i is that it is the ratio of the standard deviation of $\check{x}_{it}$ under high and low volatility regime:

$$h_i = \frac{\sigma(\check{x}_{it}\mathscr{I}(s_{it}=1))}{\sigma(\check{x}_{it}\mathscr{I}(s_{it}=0))} \text{ for } i = 1, \ldots, n \tag{19}$$

Given ψ_2, the variance of low volatility regime is known, which is given by the diagonal entries of Σ. Further condition on ψ_1 and ψ_3, the inference on h_i boils down to the inference of the variance of a normally distributed univariate random variable with known mean.

$$\check{x}_{it}\mathscr{I}(s_{it}=1) \sim \mathscr{N}(\check{\mu}_i, \check{\sigma}_i^2) \text{ for } i = 1, \ldots, n \tag{20}$$

We impose an inverse gamma prior:

$$\check{\sigma}_i^2 \sim g^{-1}\left(\frac{\nu^0}{2}, \frac{\nu^0(\sigma^0)^2}{2}\right) \text{ for } i = 1, \ldots, n \tag{21}$$

which leads to an inverse gamma posterior:

$$\check{\sigma}_i^2 \sim g^{-1}\left(\frac{\nu^1}{2}, \frac{\nu^0(\sigma^1)^2}{2}\right) \text{ for } i = 1, \ldots, n \tag{22}$$

where $\nu^1 = \nu^0 + T$, and $(\sigma^1)^2 = \frac{1}{\nu^1}(\nu^0(\sigma^0)^2 + \sum_{i\in\mathbb{B}_i}(\check{x}_{it}\mathscr{I}(s_{it}=1) - \check{\mu}_i)^2$. We denote $\mathbb{B}_i = \{i \mid s_{it} = 1\}$ as the set of high volatility observations for country i. Since our identification restriction is $h_i > 1$, we keep drawing from the posterior until such restriction is satisfied.

(4) $\boldsymbol{\psi}_4 = \{S_T\}$

The key feature of simulation-based Bayesian inference of hidden Markov model is the simulation of the states from the joint conditional distribution of all states given other parameters in the model. The procedure for drawing states is based on Albert and Chib (1993). We

avoid the intractable simulation of the whole chain at a time by drawing a single state at each step recursively. The conditional distribution that we hope to simulate from is $P(S_T \mid \boldsymbol{\psi}_{-4})$, which could be written as:

$$\begin{aligned} P(S_T \mid \boldsymbol{\psi}_{-4}, \Omega_T) = & P(s_{1T}, \ldots, s_{nT} \mid \boldsymbol{\psi}_{-4}, \Omega_T) \times \ldots \\ & \times P(s_{1t}, \ldots, s_{nt} \mid \boldsymbol{\psi}_{-4}, \Omega_T, S^{t+1}) \times \ldots \\ & \times P(s_{11}, \ldots, s_{n1} \mid \boldsymbol{\psi}_{-4}, \Omega_T, S^2) \end{aligned} \tag{23}$$

where $S_t = (s_{11}, \ldots, s_{n1}, \ldots, s_{1t}, \ldots, s_{nt})$ is the history of states up to time t, as defined earlier. And $S^{t+1} = (s_{1t+1}, \ldots, s_{nt+1}, \ldots, s_{1T}, \ldots, s_{nT})$ is the future of states from $t+1$ to T. $\Omega_t = (Y_t, X_t, Z_t)$, which is the collection of information on dependent, independent variables and exogenous drivers of the regime switching process up to time t. A typical elements in equation (21), excluding the terminal point, is

$$P(s_{1t}, \ldots, s_{nt} \mid \boldsymbol{\psi}_{-4}, \Omega_T, S^{t+1}) \tag{24}$$

By the argument in Albert and Chib (1993),

$$\begin{aligned} & P(s_{1t}, \ldots, s_{nt} \mid \boldsymbol{\psi}_{-4}, \Omega_T, S^{t+1}) \propto \\ & P(s_{1t}, \ldots, s_{nt} \mid \boldsymbol{\psi}_{-4}, \Omega_t) \times P(s_{1t+1}, \ldots, s_{nt+1} \mid s_{1t}, s_{2t}, \Omega_t, \boldsymbol{\psi}_{-4}) \end{aligned} \tag{25}$$

$P(s_{1t}, \ldots, s_{nt} \mid \boldsymbol{\psi}_{-4}, S^{t+1})$ is the product of two terms. The first term is the mass function of $(s_{1t}, \ldots, s_{nt})$ given Ω_t and other parameters in the model. This term can be derived iteratively by a prediction step and an update step. These mass functions $P(s_{1t}, \ldots, s_{nt} \mid \boldsymbol{\psi}_{-4}, \Omega_t)_{t=1}^{T}$ are stored in a $T \times 2^n$ matrix F since there are 2^n possible combinations of $(s_{1t}, \ldots, s_{nt})$ for each t. The second term is the transition probability, which can be derived given ψ_6. The last state $(s_{1T}, \ldots, s_{nT})$ is simulated using $P(s_{1T}, \ldots, s_{nT} \mid \boldsymbol{\psi}_{-4}, \Omega_T)$, which is the last row of F. And then the remaining states can be simulated using Equation 23.

(5) $\boldsymbol{\theta}_5 = \{S_T^*\}$

Performing direct inference on $\{\boldsymbol{\gamma}, \boldsymbol{\lambda}\}$ is complicated since no conjugate prior exists for the parameters of the auxiliary probit regression model. In the spirit of Kaufmann (2015), I overcome this problem by augmenting the original model in the following way:

$$s_{it}^* = \boldsymbol{z}_{it}' \boldsymbol{\gamma}_{\boldsymbol{i}} + \sum_{j \neq i} \lambda_{ij} * s_{j,t-1} + u_{it} \text{ for } i = 1, \ldots, n \tag{26}$$

Given ψ_4, ψ_6, $z'_{it}\boldsymbol{\gamma_i} + \sum_{j\neq i}\lambda_{ij} * s_{j,t-1}$ can be calculated. And u_{it} is draw from a standard normal distribution that is consistent with the prediction of the random utility model. That is, a draw u_{it} will be accepted only if

$$\begin{cases} z'_{it}\boldsymbol{\gamma_i} + \sum_{j\neq i}\lambda_{ij} * s_{j,t-1} + u_{it} \geqslant 0 \text{ if } s_{it} = 1 \\ z'_{it}\boldsymbol{\gamma_i} + \sum_{j\neq i}\lambda_{ij} * s_{j,t-1} + u_{it} < 0 \text{ if } s_{it} = 0 \text{ for } i = 1,\ldots,n \end{cases} \tag{27}$$

When we have our first accepted draw $u_{it}^{(1)}$, the latent index s_{it}^* is computed as $s_{it}^* = z'_{it}\boldsymbol{\gamma_i} + \sum_{j\neq i}\lambda_{ij} * s_{j,t-1} + u_{it}^{(1)}$

(6) $\boldsymbol{\psi}_6 = \{\boldsymbol{\gamma}, \boldsymbol{\lambda}\}$

$$s_{it}^* = \tilde{z}'_{it}\tilde{\boldsymbol{\gamma}}_i + u_{it} \text{ for } i = 1,\ldots,n \tag{28}$$

Given ψ_4 and ψ_5, we have a linear regression model, where $\tilde{\boldsymbol{z}}_{it} = (\boldsymbol{z}'_{it}, \boldsymbol{s}'_{-i})'$ and s_{it}^* are observed. $\tilde{\boldsymbol{\gamma}}_i = (\boldsymbol{\gamma}'_i, \boldsymbol{\lambda}'_i)'$. Again, I impose a conjugate normal prior:

$$\tilde{\boldsymbol{\gamma}}_i \sim \mathcal{N}(\tilde{\boldsymbol{\gamma}}_i^0, M_i^0) \text{ for } i = 1,\ldots,n \tag{29}$$

which leads to normal posterior:

$$\tilde{\boldsymbol{\gamma}}_i \sim \mathcal{N}(\tilde{\boldsymbol{\gamma}}_i^1, M_i^1) \text{ for } i = 1,\ldots,n \tag{30}$$

where $M_i^1 = ((M_i^0)^{-1} + \tilde{Z}'_i\tilde{Z}_i)^{-1}$ and $\tilde{\boldsymbol{\gamma}}_i^1 = M_i^1((M_i^0)^{-1}\tilde{\boldsymbol{\gamma}}_i^0 + \tilde{Z}'_i S^*)$

Chapter 2

News-Implied Linkages and Local Dependency in the Equity Market

Abstract

This paper studies a heterogeneous coefficient spatial factor model that separately addresses both common factor risks (strong cross-sectional dependence) and local dependency (weak cross-sectional dependence) in the equity returns. From the asset pricing perspective, we derive the theoretical implications of no asymptotic arbitrage for the heterogeneous spatial factor model. In empirical work, it is challenging to measure granular firm-to-firm connectivity for a high-dimensional panel of equity returns. We use extensive business news to construct firms' links via which local shocks transmit, and we use those news-implied linkages as a proxy for the connectivity among firms. Empirically, we document a considerable degree of local dependency among *S&P* 500 stocks, and the spatial component does a great job in capturing the remaining correlations in the de-factored returns. We find that adding spatial interactions to factor models reduces mispricing and mean-squared errors. We also show that our news-implied linkages provide a comprehensive and integrated proxy for firm-to-firm connectivity, and it out-performs other existing networks in the literature.

Keywords: Spatial asset pricing model; weak and strong cross-sectional dependence; local dependency; networks; textual analysis; big data; large heterogeneous panel
JEL Classification: C33; C58; G10, G12

2.1 Introduction

Comovement in equity returns are the combined effects of exposures to common risks and local interactions. Classical asset pricing models such as the classical capital asset pricing model (CAPM) developed by Sharpe (1964), and the arbitrage pricing theory (APT) by Ross (1976) focus only on the strong pervasive component driven by a few common factors. Many studies have found those models focusing on strong dependence only are insufficient to capture all the significant interdependencies in asset returns. Local dependencies still play a non-negligible role (see for example Gabaix (2011), Acemoglu et al. (2012), Israelsen (2016), Barigozzi and Hallin (2017), Kou et al. (2018), Hale and Lopez (2019), Bailey et al. (2019a), and Barigozzi and Brownlees (2019)).

In this paper, we study a spatial factor model that separately addresses both common factor risks and local dependencies in the equity returns. The factor component and spatial component complement each other, with the former capturing strong cross-sectional dependence in equity returns and the latter capturing the weak cross-sectional dependence due to local interactions among entities. To distinguish the two sources of dependencies, imagine a group of people sitting in a room on a chilly winter day. People might catch a cold because the heater is broken (common factors) or because someone sitting close to them is ill (local interactions). The network architecture of entities, like the sitting plan in the previous example, is key to studying local interactions. Unlike spatial interactions in geographical systems, where there exists a natural network structure, for a high dimensional panel of equity returns, there is no natural network structure.

We use extensive business news data to construct firms' linkages via which local shocks transmit, and we use those news-implied linkages as a proxy for the connectivity among firms. It has been documented that common news coverage reveals information about linkages among companies, which are related to many economically important relationships like business alliances, partnerships, banking and financing, customer-supplier, and production similarity (Scherbina and Schlusche (2015), Schwenkler and Zheng (2019)). We use news data from RavenPack Equity files Dow Jones Edition for the period between the beginning of 2004 to the end of 2015. This comprehensive news dataset combines relevant content from multiple sources, including Dow Jones Newswires, Wall Street Journal, and Barron's MarketWatch, which produce the most actively monitored streams of news articles in the financial system. We identify linkages among firms by news co-mentioning.

Using the novel text-based network, we estimate spatial factor models with different sets of common risk factors. We find a considerable degree of local dependencies among *S&P*

500 stocks. The spatial interaction terms are highly significant after controlling for popular factors, and they continue to be significant even after adding industry-level factors. Different from most spatial econometrics literature, where spatial coefficients are assumed to be homogeneous, we adopt a heterogeneous coefficient framework from Bailey et al. (2016) and Aquaro et al. (2020). The model is very flexible, allowing us to capture general local interactions pattern among a large number of firms. Using that framework, we are able to not only investigate the average effect among all or some subgroups of firms but also gauge the individual-level effect. We document that apart from the average spatial effect measured by the mean group (MG) estimator being highly significant, at the individual level, spatial effect along news-implied linkages are also highly significant. We find that the percentage of individual contemporaneous spatial parameter being statistically significant at 5% level is over 88% across all specifications that we consider. This high significance ratio implies that the news-based link identification method is successful at detecting economically important links. The framework also allows us to examine the heterogeneity at subgroup levels. By applying mean group estimation to different industry groups, we document heterogeneity at the industry level. In particular, financial companies have the highest degree of local dependencies. We argue that the spatial factor model provides a unified way of addressing both strong and weak/local dependence in the equity returns. To investigate how well the spatial factor model captures the remaining dependence in the de-factored component, we examine the changes in correlation structure before and after adding the spatial component to the traditional factor models. We find that adding the spatial component reduces the number of non-zero pair-wise cross correlations by a huge margin, and the spatial factor model error correlation matrix is very close to diagonal. These results show that the spatial component constructed with news-implied linkages is successful at eliminating remaining correlations from the de-factored returns. We also compare the degree of mispricing and mean-squared errors for a set of factor models and their spatial augmented versions. We find that adding spatial/local interaction terms significantly reduces mispricing and mean-squared errors.

This paper contributes to three strands of literature. The first one is cross-sectional dependencies in equity returns. Cross-sectional dependence in a large panel is usually complex and reflects different types of interdependencies. Chudik et al. (2011), and Bailey et al. (2016) show that strong cross-sectional dependence (CSD) and weak cross-sectional dependence (CWD) have different economic implications and statistical behaviors, thus need to be accounted for separately. Kuersteiner and Prucha (2020) consider a short T panel with cross-sectional dependence due to both common factor risks and spatial/local interactions. While asset pricing literature has been focused on strong dependence (i.e., exposures to common risk factors), local dependence receives much less attention theoretically and empirically.

Theoretically, we extend classical arbitrage pricing theory (APT), which only take strong cross-sectional dependence into account. We propose a flexible spatial factor model that addresses both strong and weak/local dependence for a large panel in a single framework. Especially, we derive the implications of no asymptotic arbitrage for our heterogeneous coefficient spatial factor model. Empirically, we show the benefit from addressing spatial/local interaction in terms of reducing mispricing errors and mean-squared errors.

Another major contribution is that we use a novel way to proxy the local connectivity among a large set of firms. Fan et al. (2011) suppose that the error covariance matrix is sparse (i.e., has lots of zeros), which represents the absence of linkages between firms beyond that contained in the common factors. They identify the location of non-zero entries by applying thresholding methods to the error sample covariance matrix. Our method uses information gathered from other sources, specifically news stories, to identify the linkages. There has been exploding research on quantifying the information embeded in unstructured data like text data. Alternative data fill the gaps in data availability induced by limited disclosure and slow update, thus complementing traditional economic datasets. For example, there has been a steep rise in the number of studies on utilizing the information from text (Garcia (2013), Scherbina and Schlusche (2015), Baker et al. (2016), Hoberg and Phillips (2016), Ke et al. (2019), Schwenkler and Zheng (2019), etc). This paper explores a comprehensive news dataset that combines relevant content from multiple sources and identifies linkages among firms by news co-mentioning. With a measure of local connectivity, we can capture correlations from both strong and the remaining weak dependence in a large panel using a single step. Without a knowledge of local connectivity, Fan et al. (2011) need a two-step procedure (they first estimate a factor model, and then use thresholding to estimate the error sample covariance matrix).

Our work also contributes to network effect or local risk spillover effect among economically linked firms. Local risks transmit through economic linkages, and firms with links exhibit excess co-movement. There has been various proxies for firm to firm networks in the literature, including industry-based peers (Moskowitz and Grinblatt (1999), Fan et al. (2016a), and Engelberg et al. (2018)), analyst co-coverage networks (Kaustia and Rantala (2013), Israelsen (2016), and Ali and Hirshleifer (2020)), customer-supplier networks (Cohen and Frazzini (2008)), geographic networks (Pirinsky and Wang (2006), Parsons et al. (2020)), etc. We show that our news-based linkages provide a comprehensive and integrated proxy for firm-level connectivity. Spatial factor model estimated with news-implied network outperforms those aforementioned networks in terms of minimizing the mispricing errors and the mean-squared errors. Even if we consider the union of all those competing networks,

news-implied networks provide equally good performance. We also show that conditional on all those previously documented links, our news-implied linkages are still important channels of local risk spillovers.

The rest of the paper is organized as follows. Section 2 describes the difference between strong and weak cross-sectional dependence and introduces the spatial factor model. Section 3 develops the asset pricing implications with the presence of local interactions. Section 4 shows the estimation and inference of the heterogeneous coefficient spatial-temporal model that we use. Section 5 presents the estimation results, performances of alternative models, and comparisons with previously documented networks. Section 6 concludes. Proofs, technical details, and supplementary figures and tables are in the Appendices.

Notations: If $\{f_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$ are both positive sequence of real numbers, then $f_n = \Theta(g_n)$ if there exist $N_0 \geqslant 1$ and positive finite constants C_0 and C_1, such that $inf_{n \geqslant N_0}(f_n/g_n) \geqslant C_0$, and $sup_{n \geqslant N_0}(f_n/g_n) \leqslant C_1$. For a $N \times N$ real matrix $A = (a_{ij})$, define its maximum column sum norm by $\|A\|_1 = \max_{1 \leqslant j \leqslant N} \sum_{i=1}^{N} \mid a_{ij} \mid$, and its maximum row sum norm by $\|A\|_\infty = \max_{1 \leqslant i \leqslant N} \sum_{j=1}^{N} \mid a_{ij} \mid$.

2.2 Modelling Cross-Sectional Dependence by Spatial Factor Model

2.2.1 Strong Dependence: Factor Model

Consider the below factor model

$$\boldsymbol{r}_t - r_{ft}\mathbf{1} = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\varepsilon}_t, t = 1, \ldots, T, \tag{2.1}$$

where $\boldsymbol{r}_t$ is the $N \times 1$ vector of equity returns at t, r_{ft} is the risk free rate at t, and $\mathbf{1}$ is $N \times 1$ vector of ones. $\mathbf{f}_t$ is the $K \times 1$ vector of common risk factors, and B is the $N \times K$ factor loadings, where β_{ik} is the loading of asset i on factor k. Let

$$\begin{aligned} &\sum_{i=1}^{N} \mid \beta_{ik} \mid = \Theta(N^{\alpha_{\beta_k}}), \text{ for } k = 1, \ldots, K, \\ &\|\mathbf{B}\|_1 = \max_{1 \leqslant k \leqslant K} \sum_{i=1}^{N} \mid \beta_{ik} \mid = \Theta(N^{\alpha_{\mathbf{B}}}), \end{aligned} \tag{2.2}$$

where Θ denotes the exact order of magnitude, α_{β_k} measures how pervasive the kth factor is, and $\alpha_{\mathbf{B}} = \max_k(\alpha_{\beta_k})$ measures how pervasive the factor component $\mathbf{Bf}_t$ is. In the standard factor models, it is assumed that $\alpha_{\beta_k} > 0$ for $k = 1,\ldots,K$, and $\alpha_{\mathbf{B}} = 1$, so that each factor has non-diminishing effect on the system and exposures to common risk factors give rise to strong cross-sectional dependence, which is systematic and non-diversifiable.

2.2.2 Weak/Local Dependence: Spatial Model

Consider the below canonical spatial autoregressive model with homogeneous spatial coefficient

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \psi W \boldsymbol{r}_t + \boldsymbol{\varepsilon}_t, t = 1,\ldots,T, \tag{2.3}$$

where W is the $N \times N$ adjacency matrix that specifies the channels from which shocks transmit, where a typical entry w_{ij} gives the influence of the returns of j on that of i. The strength of spatial risk spillovers is represented by the parameter ψ.

Spatial dependence characterizes weak cross-sectional dependence where interactions are local. For demonstration, we re-write Equation 2.3 as follows:

$$\boldsymbol{r}_t = G(\psi)\boldsymbol{\nu}_t, \tag{2.4}$$

where $G(\psi) = (I_N - \psi W)^{-1}$ and $\boldsymbol{\nu}_t = \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_t$. Equation 2.4 can be interpreted as a factor model with N factors. $G(\psi)$ is the $N \times N$ factor loadings, where g_{ij} is the loading of i on factor j. All factors are weak and only have local effects if the following absolute summability condition is true

$$\sum_{i=1}^{N} \mid g_{ij} \mid \leqslant c \quad \text{for } j = 1,\ldots,N, \text{ where } c \text{ is a positive constant.} \tag{2.5}$$

The absolute summability condition Equation 2.5 is equivalent to a bounded column sum matrix norm condition on the Leontief inverse $G(\psi) = (I_N - \psi W)^{-1}$. As in LeSage (2008),

$$G(\psi) = (I_N - \psi W)^{-1} = I + \psi W + \psi^2 W^2 + \ldots = I + \sum_{j=1}^{\infty} \psi^j W^j. \tag{2.6}$$

The Leontief inverse take accounts of direct interaction effect and higher-order indirect effects. The assumption that the column sum of $G(\psi) = (I_N - \psi W)^{-1}$ is uniformly bounded in the number of cross-sectional units N is usually assumed in early spatial econometrics (see Kelejian and Prucha (1998), Kelejian and Prucha (1999)) to limit the cross-sectional

correlation to a manageable degree. Although more recent development shows we do not need such assumption to guarantee stationary of the model and consistent estimation (see Elhorst et al. (2014), Aquaro et al. (2020)), we take that assumption as a modelling assumption to distinguish strong and weak/local dependence. In particular, for weak dependence, no cross-sectional unit exerts pervasive effects on the system and the interactions are local. There will be more discussions in subsection 2.2.3.

2.2.3 Strong and Weak Dependence: Spatial Factor Model

Comovement in a large panel of equity returns arises due to both strong and weak cross-sectional dependence. Many studies have found that factor models that only focus on the non-diversifiable type of risks are insufficient to capture all the cross-sectional dependence in equity returns. We study a heterogeneous coefficient spatial factor model written as Equation 2.7 where the factor component and the spatial component complement each other, with the former addressing strong dependence and the latter addressing spillovers that are non-pervasive/local in nature (i.e., cross-sectional weak dependence (CWD) define in Chudik et al. (2011)).

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \Psi W \boldsymbol{r}_t + \boldsymbol{\varepsilon}_t, t = 1, \ldots, T, \tag{2.7}$$

where $\Psi = diag(\boldsymbol{\psi}) = diag(\psi_1, \ldots, \psi_N)$ is a diagonal matrix with N individual specific contemporaneous spatial coefficients on the main diagonals.

The spatial component has several main features. Firstly, the spillover coefficients are heterogeneous. One might reasonably suspect that the sensitivities to neighbours' risks are different from firm to firm. While the restrictive assumption that all entities share the same spatial coefficient is necessary for small T, it can be relaxed when T is big. Aquaro et al. (2020) shows that a heterogeneous spatial autoregressive model like Equation 2.7 can be consistently estimated with large T. We utilize this nice feature to explore the heterogeneity in the strength of local dependency. Moreover, we could examine the heterogeneity pattern at subgroup levels (such as industry levels) using mean-group estimation, which is a popular tool in heterogeneous panel literature. Secondly, it is possible to add weakly exogenous spatial-temporal terms $\sum_{l=1}^{L} \Psi_l W \boldsymbol{r}_{t-l}$ to Equation 2.7. $\Psi_l W \boldsymbol{r}_{t-l}$ corresponds to the spatial-temporal term at the lth lag for $l = 1, \ldots, L$, where $\Psi_l = \text{diag}(\psi_l) = \text{diag}(\psi_{l,1}, \ldots, \psi_{l,N})$ is a diagonal matrix of spatial-temporal parameters at the lth lag. These dynamic terms may account for potential market microstructure effects, which is important in our empirical application to daily individual stock returns.

Kou et al. (2018) consider a special case of this model, which they call the Spatial APT model. In particular, they consider the case where N is small, $L = 0$ (no temporal dynamics), $\psi_i = \psi$ (homogeneous spatial effects), and homoscedastic errors. Kou et al. (2018) derives the implications of the absence of arbitrage on the parameters of the model, in particular on the intercept vector $\boldsymbol{\alpha}$. In section 2.3, we extend their analysis by deriving the implications of no arbitrage under our framework.

In this paper, we impose the following assumptions on the spatial factor model (Equation 2.7):

Assumption 1. $E(\boldsymbol{\varepsilon}_t) = 0,\ E(\mathbf{f}_t \boldsymbol{\varepsilon}_t') = 0,\ E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \Omega,\ \sigma_i^2 = var(\varepsilon_{it}) \leqslant \bar{\sigma}^2.$

Assumption 2. *Let* $\sum_{i=1}^{N} \mid \beta_{ik} \mid = \Theta(N^{\alpha_{\beta_k}}),$ *for* $k = 1, \ldots, K$, *and* $\|\mathbf{B}\|_1 = \max_{1 \leqslant k \leqslant N} \sum_{i=1}^{N} \mid \beta_{ik} \mid = \Theta(N^{\alpha_{\mathbf{B}}})$ *as Equation 2.2.* $\alpha_{\beta_k} > 0$ *for* $k = 1, \ldots, K,$ *and* $\alpha_{\mathbf{B}} = 1.$

Assumption 3. $\sup_i \mid \psi_i \mid < \frac{1}{\|W\|_\infty},$ *for* $i = 1, \ldots, N.$

Assumption 4. *Define* $G(\Psi) = (I_N - \Psi W)^{-1}$, *where* g_{ij} *is a typical entry of* G. $\sum_{i=1}^{N} \mid g_{ij} \mid \leqslant c$ *for* $j = 1, \ldots, N$ *for some positive constant* c.

Remark 1: Assumption 2 guarantees that each factor has non-diminishing effect on the system, and the exposures to common risk factors give rise to strong cross-sectional dependence. We only assume that at least one factor is strong (i.e., with $\alpha = 1$), and all other factors are not weak. This is because in practice we may want to add many factors, which have different degrees of pervasiveness. For example, Bailey et al. (2020) find that for the factors proposed in the finance literature for asset pricing, only the market factor is strong over all the windows they consider.
Remark 2: Assumption 3 is to ensure $(I_N - \Psi W)$ is invertible and $G(\Psi) = (I_N - \Psi W)^{-1}$ exists.
Remark 3: Assumption 4 assumes that the column sums of $G(\Psi) = (I_N - \Psi W)^{-1}$ is uniformly bounded in absolute values as N goes to infinity. This ensures that no cross-sectional unit exerts pervasive effects on the system and the interactions are local. This is a modelling assumption, which is not required for stationarity and consistent estimation of the model. From our view, the correlation beyond factors should be weak. Pervasive dependence should be addressed by adding sufficient common factors into the model. Similar assumptions are made in Fan et al. (2008) and Fan et al. (2011), where they assume that after taking out the influence of Fama-French three factors, the remaining cross-sectional dependence is weak in the way defined in Chamberlain and Rothschild (1983).

The spatial factor model provides a unified way of addressing the remaining dependence in the de-factored component. Fan et al. (2011) identify the location of significant correlations by applying thresholding methods to the factor model error sample covariance matrix. To capture both factor-driven strong dependence and remaining weak dependence in a large panel, they need a two-step procedure. Our method provides an alternative, which can be achieved in a single step. Compared with purely statistical methods, our method also has the advantage of being interpretable given that our linkages are constructed using information from business news.

To investigate how well the spatial factor model captures the remaining dependence in the de-factored component, we can examine the changes in correlation structure before and after adding the spatial component to factor models. If the spatial component is doing a good job in terms of explaining the remaining local dependence, then we should expect to see the number of pairs with non-zero pair-wise error cross correlations being reduced by adding the spatial component. In our application, we estimate the number of non-zero pair-wise cross correlations of residuals from (1) a set of factor models, and (2) their corresponding spatial augmented models. For N cross-sectional units, the problem considers testing $N(N-1)/2$ null hypotheses simultaneously. We use multiple testing procedure to control for the overall size of the tests.

Under the factor model settings (Equation 2.1), this task is relatively easy. Pesaran et al. (2004) establishes the asymptotic distribution of error correlation coefficient under the null $H_{0,ij} : \rho_{ij} = 0$ for panel data models as follows:

$$y_{it} = \alpha_i + \boldsymbol{\beta}_i' \boldsymbol{x}_{it} + \varepsilon_{it}, \tag{2.8}$$

where $Var(\boldsymbol{\varepsilon}_t) = \Sigma = (\sigma_{ij})$ is an $N \times N$ symmetric, positive definite matrix. Denote the correlation coefficient of ε_{it} and ε_{jt} by ρ_{ij}. To estimate the correlation coefficient of errors, one needs to first obtain residuals $\hat{\varepsilon}_{it}$ as

$$\hat{\varepsilon}_{it} = y_{it} - \boldsymbol{x}_{it}'(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{y}_i, \tag{2.9}$$

where $\boldsymbol{X}_i$ is the $T \times K$ matrix of regressors for unit i, and $\boldsymbol{y}_i$ is the $T \times 1$ vector of the dependent variable for unit i. The sample estimate of ρ_{ij} is given by

$$\hat{\rho}_{ij} = \frac{\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_j/T}{(\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i/T)^{1/2}(\hat{\boldsymbol{\varepsilon}}_j'\hat{\boldsymbol{\varepsilon}}_j/T)^{1/2}} \tag{2.10}$$

When $\boldsymbol{x}_{it}$ are strictly exogenous, under the null $H_{0,ij}: \rho_{ij} = 0$,

$$\sqrt{T}\hat{\rho}_{ij} \to N(0,1) \text{ as } T \to \infty. \tag{2.11}$$

To test the hypothesis $H_{0,ij}: \rho_{ij} = 0$, for p being the chosen nominal size, we can use $\frac{1}{\sqrt{T}}\Phi^{-1}(1-p/2)$ as the critical value, where Φ^{-1} is the inverse cdf of standard normal. However, to test $\rho_{ij} = 0$ for all $i \neq j$ jointly, we need to take the multiple testing issue into account. From Bonferroni (1935), given that there are $n_{test} = N(N-1)/2$ such tests, for the family-wise error rate (FWER) to be p, it is sufficient to set the nominal size for each individual test $p_i = p/n_{test}$ for $i = 1,\ldots,n_{test}$, so that the critical value for each $\mid \hat{\rho}_{ij} \mid$ becomes $\frac{1}{\sqrt{T}}\Phi^{-1}(1-p/2n_{test})$.[1]

The nice theoretical result (Equation 2.11) is derived under the assumptions of exogeneity of regressors. However, the spatial factor model this is not the case. For the below spatial factor model we consider (Equation 2.7), the spatial autoregressive term $W\boldsymbol{r}_t$ is endogenous, which makes the result from Equation 2.11 fail. Given that, we conduct bootstrap inference for the error correlations of the spatial factor model, which proceeds in the following steps.

We first estimate the spatial factor model (Equation 2.7) and collect the estimated parameters $\hat{\boldsymbol{\alpha}}, \hat{\mathbf{B}}, \hat{\Psi}$. We save the residual values $R = \{\hat{\varepsilon}_{11},\ldots,\hat{\varepsilon}_{1T},\ldots,\hat{\varepsilon}_{N1},\ldots,\hat{\varepsilon}_{NT}\}$. In the next step, for $b = 1,\ldots,B$, under the null of diagonal error correlation matrix, we draw $i.i.d$ $\hat{\varepsilon}_{it}^b$ from R, and generate the bth bootstrap sample as

$$\boldsymbol{r}_t^b = (I - \hat{\Psi}W)^{-1}(\hat{\boldsymbol{\alpha}} + \hat{\mathbf{B}}\mathbf{f}_t + \boldsymbol{\varepsilon}_t^b), t = 1,\ldots,T. \tag{2.12}$$

We re-estimate the model using the bootstrap sample. Next, we calculate the sample correlation coefficients $\hat{\rho}_{ij}^b$ for all $i \neq j$. We save those $N(N-1)/2$ pair-wise cross-correlations for each bootstrap sample b. Finally, we can draw inference from the empirical null distribution F by computing the critical values associated with a nominal size value p as $F^{-1}(p/2)$ and $F^{-1}(1-p/2)$. Again, we need to correct for multiple testing issue here, and the critical values for each $\hat{\rho}_{ij}$ becomes $F^{-1}(p/2n_{test})$ and $F^{-1}(1-p/2n_{test})$.

[1] There are more advanced methods of choosing threshold values, like Bailey et al. (2019b). However, the theory does not go through for testing error correlation of the spatial factor model. To have a fair comparison, we consider Bonferroni type of correction for both factor and spatial factor model.

2.3 Arbitrage Pricing Theory Under Spatial Factor Model

In this section, we derive the asset pricing implications of our heterogeneous coefficient spaital factor model (Equation 2.7). We follow Ingersoll Jr (1984), and consider a fixed infinite economy where a sequence of nested subsets of assets are examined. For the nth economy, a new asset is added to the $(n-1)$th economy.

$$\boldsymbol{r}^{n\prime} = (\boldsymbol{r}^{n-1\prime}, r^n). \tag{2.13}$$

Since APT is a cross-sectional approach, we drop the time subscript. We denote the size of economy by superscript n, and a portfolio in the nth economy is denoted as $\boldsymbol{c}^n \in R^n$. $\mathbf{1}^n$ is the vector of n ones. We consider subsequences of assets, where subsequences are indexed by v. There are asymptotic arbitrage opportunities if there is a subsequence of portfolios that satisfy the following conditions:

$$\begin{aligned} &Var(\boldsymbol{c}^{v\prime}\boldsymbol{r}^{\boldsymbol{v}}) \to 0 \ \ as \ \ \ v \to \infty, \\ &E(\boldsymbol{c}^{v\prime}\boldsymbol{r}^{\boldsymbol{v}}) \geqslant \delta > 0 \ \ \text{for all } v, \\ &\boldsymbol{c}^{v\prime}\mathbf{1}^v = 0 \ \ \text{for all } v. \end{aligned} \tag{2.14}$$

Theorem 1. *Assume that the returns are generated by the heterogeneous spatial factor model (Equation 2.7), and Assumption 1-4 hold. If there is no arbitrage opportunities described in Equation 2.14, then there is a sequence of K by 1 vector of factor premiums $\boldsymbol{\lambda}^n$ and a constant λ_0^n such that the following approximation holds*

$$\boldsymbol{\alpha}^n \approx (I_n - \Psi^n W^n)\mathbf{1}^n\lambda_0^n + \mathbf{B}^n\boldsymbol{\lambda}^n. \tag{2.15}$$

Define pricing error vector as:

$$\boldsymbol{v}^n = \boldsymbol{\alpha}^n - (I_n - \Psi^n W^n)\mathbf{1}^n\lambda_0^n - \mathbf{B}^n\boldsymbol{\lambda}^n. \tag{2.16}$$

The approximation Equation 2.15 holds in a sense that there is a positive number V such that the weighted sum of squared pricing errors is uniformly bounded,

$$(\boldsymbol{v}^n)'(\Omega^n)^{-1}(\boldsymbol{v}^n) \leqslant V < \infty \ \ \textit{for all } n. \tag{2.17}$$

Corollary 1.1. *Let λ_{max} denote the largest eigenvalue of the limit covariance matrix Ω^n, then $V = q\lambda_{max}$ for a positive number q.*

Corollary 1.2. *Suppose all factors are traded. The risk factor f_k before de-meaning is denoted as $\tilde{f}_k$, which is the payoff of the kth tradable zero-cost portfolio, where*

$$f_k = \tilde{f}_k - E(\tilde{f}_k) \text{ for } k = 1, \ldots, K. \tag{2.18}$$

If there exists a risk free asset with rate r_f, the spatial factor model (Equation 2.7) can be written as:

$$\begin{aligned} &\tilde{\boldsymbol{r}} = \tilde{\boldsymbol{\alpha}} + \mathbf{B}\tilde{\mathbf{f}} + \Psi W \tilde{\boldsymbol{r}} + \boldsymbol{\varepsilon}, \\ &\text{where } \tilde{\boldsymbol{r}} = \boldsymbol{r} - r_f \mathbf{1} \text{ is the vector of excess returns,} \\ &\tilde{\mathbf{f}} = (\tilde{f}_1, \ldots, \tilde{f}_K)', \\ &\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} - (I - \Psi W)\mathbf{1} r_f - \mathbf{B}E(\tilde{\mathbf{f}}). \end{aligned} \tag{2.19}$$

Then the no arbitrage condition for an infinite economy where asset returns are generated by the spatial factor model is:

$$\tilde{\boldsymbol{\alpha}}^n \approx 0. \tag{2.20}$$

$\tilde{\boldsymbol{\alpha}}^n$ is the pricing error in this special case. The approximation Equation 2.20 holds in a sense that there is a positive number V such that the sum of squared pricing errors is uniformly bounded,

$$(\tilde{\boldsymbol{\alpha}}^n)'(\Omega^n)^{-1}(\tilde{\boldsymbol{\alpha}}^n) \leqslant V < \infty \text{ for all } n. \tag{2.21}$$

And λ_0^n, $\boldsymbol{\lambda}^n$ in 1 can be identified as

$$\begin{aligned} \lambda_0^n &= r_f, \\ \boldsymbol{\lambda}^n &= E(\tilde{\mathbf{f}}). \end{aligned} \tag{2.22}$$

Corollary 1.3. *In addition to the assumptions from Corollary 1.2 (there exists risk free assets and all factor are traded), if we further assume that errors are uncorrelated, then for any $\delta > 0$, there is a constant N_δ such that the number of elements in $\tilde{\boldsymbol{\alpha}}$ that are bigger than δ in absolute values is uniformly bounded by N_δ,*

$$\lim_{n\to\infty} \sum_{j=1}^{n} I(\mid \tilde{\alpha}_j \mid > \delta) < N_\delta < \infty \tag{2.23}$$

Remark 1: Corollary 1.1 implies that the correlation structure of $\Omega = E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')$ affects the bound V. The less correlation there in Ω, the smaller V is, and the better the approximation implied by Equation 2.15 is. Spatial factor model addresses weak/local dependence beyond strong dependence captured by factors, and we expect that there is less correlation in

Ω. Compared with factor models, the spatial factor model is expected to give a better approximation.

Remark 2: Corollary 1.2 points out a special case useful for empirical work. Assuming that there exists risk free asset and all factors are traded, we can just look at the spatial factor model with the dependent variable being the excess returns (Equation 2.19). In particular, to test the theoretical restrictions we just need to examine how close to zero the intercept vector $\tilde{\boldsymbol{\alpha}}$ is.

Remark 3: Theorem 1 and the corollaries suggest some statistics that we can employ to compare the relative performance of different asset pricing models. In particular, the L_1, L_2 norms of mispricing errors, and the number of components with big mispricing errors could be useful in measuring how well the approximation is.

Remark 4: Theorem 1 and the corollaries can be easily extended to spatial factor models with more than one spatial spillover channels, for example, the two-W model in Bailey et al. (2016).

Proofs of Theorem 1, Corollary 1.1, Corollary 1.2, and Corollary 1.3 are in the section A of the Appendix.

2.4 Estimation and Inference

The Corollary 1.2 points out a special case useful for empirical work. From now on we assume there is risk free asset, and all factors are traded, and we work with the panel spatial factor model where the dependent variable is the excess returns (multi-period Equation 2.19).

$$\tilde{\boldsymbol{r}}_t = \tilde{\boldsymbol{\alpha}} + \mathbf{B}\tilde{\mathbf{f}}_t + \Psi W\tilde{\boldsymbol{r}}_t + \boldsymbol{\varepsilon}_t, t = 1, \ldots, T, \tag{2.24}$$

where $\tilde{\boldsymbol{r}}_t = (\tilde{r}_{1t}, \ldots, \tilde{r}_{Nt})$ is the N by 1 vector of excess returns at t, and $\tilde{\boldsymbol{f}}_{\boldsymbol{t}} = (\tilde{f}_{1t}, \ldots, \tilde{f}_{Kt})'$ is the vectors of K factors at t. W is the N by N adjacency matrix, where a typical entry is denoted as w_{ij}. Without loss of generality , we set $w_{ii} = 0$ for all i, and we assume that $w_{ij} \geqslant 0$. Following the convention in spatial econometrics, we further row normalize W so that $\sum_{j=1}^{N} w_{ij} = 1$ for all i. Here, $\boldsymbol{\varepsilon}_t$ is the vector of errors at t, which satisfies $\text{var}(\boldsymbol{\varepsilon}) = \Sigma = \text{diag}(\boldsymbol{\sigma}_{\varepsilon^2}) = \text{diag}(\sigma_1^2, \ldots, \sigma_N^2)$, where $\sigma_i^2 = \text{var}(\varepsilon_{it})$ is allowed to be heteroskedastic.

An extension of Equation 2.24 is to incorporate weakly exogenous spatial-temporal terms:

$$\tilde{\boldsymbol{r}}_t = \tilde{\boldsymbol{\alpha}} + \mathbf{B}\tilde{\mathbf{f}}_t + \Psi_0 W\tilde{\boldsymbol{r}}_t + \sum_{l=1}^{L} \Psi_l W\tilde{\boldsymbol{r}}_{t-l} + \boldsymbol{\varepsilon}_t, t = 1, \ldots, T, \tag{2.25}$$

where we denote the contemporaneous spatial coefficients using Ψ_0, and L is the number of spatial-temporal terms to incorporate. In what follows, we set $L = 5$ to control for within-week dynamics. These dynamic terms may account for potential market microstructure effects, which is important in our empirical application to daily individual stock returns. This modification has been used in Eugene (1992), see also Dimson (1979). We estimate the heterogeneous coefficient spatial-temporal model (Equation 2.25) by the QML procedure proposed in Bailey et al. (2016) and Aquaro et al. (2020). We collect all the parameters in Equation 2.25 in the $(N*(K+L+3))$ by 1 vector $\boldsymbol{\theta} = (\tilde{\boldsymbol{\alpha}}', \boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_K', \boldsymbol{\psi}_0', \ldots, \boldsymbol{\psi}_L', \boldsymbol{\sigma}_{\varepsilon^2}')'$, and denote the vector of true values by $\boldsymbol{\theta}_0$. The log-likelihood function of Equation 2.25 is written as follows:

$$\mathscr{L}_T(\boldsymbol{\theta}) = -\frac{NT}{2}\ln(2\pi) - \frac{T}{2}\sum_{i}^{N}\ln(\sigma_i^2) + \frac{T}{2}\ln|\boldsymbol{S}'(\boldsymbol{\psi}_0)\boldsymbol{S}(\boldsymbol{\psi}_0)| - \frac{1}{2}\sum_{t=1}^{T}[\boldsymbol{S}(\boldsymbol{\psi}_0)\tilde{\boldsymbol{r}}_t - \boldsymbol{B}\boldsymbol{x}_t]'^{-1}[\boldsymbol{S}(\boldsymbol{\psi}_0)\tilde{\boldsymbol{r}}_t - \boldsymbol{B}\boldsymbol{x}_t], \tag{2.26}$$

where $\boldsymbol{S}(\boldsymbol{\psi}_0) = I_N - \Psi_0 W$, and $\tilde{\boldsymbol{r}}_t = (\tilde{r}_{1t}, \ldots, \tilde{r}_{Nt})$. We stack the constant and all the weakly exogenous variables for i at t in $x_{it} = (1, \tilde{f}_{1t}, \ldots, \tilde{f}_{Kt}, \sum_{j=1}^{N} w_{ij}\tilde{r}_{jt-1}, \ldots, \sum_{j=1}^{N} w_{ij}\tilde{r}_{jt-L})$, and $\boldsymbol{x}_t = (x_{1t}', \ldots, x_{Nt}')'$ is the $((1+K+L)*N)$ by 1 vector. $\boldsymbol{B}$ is the N by $((1+K+L)*N)$ block diagonal matrix with elements $\boldsymbol{b}_i' = (\tilde{\alpha}_i, \beta_{1,i}, \ldots, \beta_{K,i}, \psi_{1,i}, \ldots, \psi_{L,i})'$ on the main diagonal and zeros elsewhere.

The quasi maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{QMLE}$ maximizes Equation 2.26. The error terms need not be Gaussian, but when they are, $\hat{\boldsymbol{\theta}}_{QMLE}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$. Note that conditional on $\boldsymbol{\psi}_0$, the system is linear, so that we can concentrate out the parameters $\boldsymbol{B}, \boldsymbol{\sigma}_{\varepsilon^2}$ to reduce the dimensionality and hence computational burden.

Aquaro et al. (2020) provides sufficient conditions for consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_{QMLE}$ in the case where T is large and N is fixed, and consistency and asymptotic normality for the mean group estimators in the case where both T and N are large but $\sqrt{N}/T \to 0$. Further details regarding the identification conditions and inference can be found in the section B of the appendix.

2.5 Data

We consider daily returns of *S&P* 500 stocks for our application. All the stock market related data are from the Center for Research in Security Prices (CRSP). Daily factor returns and industry classification information are obtained from Kenneth French's website. Accounting data are from the merged CRSP/Compustat database. Data used to construct alternative networks are described in details in subsection 2.6.3.

The news data are obtained from RavenPack Equity files Dow Jones Edition for the period January 2004 to December 2015. This comprehensive news dataset combines relevant content from multiple sources, including Dow Jones Newswires, Wall Street Journal, and Barron's MarketWatch, which produce the most actively monitored streams of news articles in the financial system. Each unique news story (identified by unique story ID) tags the companies mentioned in the news by their unique and permanent entity identifier codes (RP_ENTITY_ID), by which we link to stock identifier TICKER and PERMNO.

Inspired by Scherbina and Schlusche (2015) and Schwenkler and Zheng (2019), we identify links by news co-mentioning. To be more specific, if a piece of business news reports two and only two companies together, then the two firms have some business relationship/link. Although news that mentions more than two companies together may carry potential information about links, they provide noisier information. We also remove news with topics including analyst recommendations, rating changes, and index movements as these types of news might stack multiple companies together when they actually do not have real links. Table 10 provides descriptive statistics for RavenPack Equity files Dow Jones Edition dataset during the sample period. Since our comprehensive news dataset combines several sources, given a similar length of sample period, the number of unique news stories is more than ten times larger than that from Scherbina and Schlusche (2015) and more than eight hundred times than that from Schwenkler and Zheng (2019). For link identification purposes, we only use sample news (1) are not about topics mentioned above, (2) tag *S&P* 500 companies, and (3) mention exactly two companies, which is a subsample of $1,637,256$ unique news stories.

From all the links identified using this methodology, some links are transitory while some are more long-lasting. To gauge the persistency of links, we split full sample news data into 12 yearly link identification windows. Table 11 is the frequency distribution table of the number of yearly link identification windows that a pair gets identified as economic neighbours for all possible pairs (i, j) in our sample. 72.80% of the pairs never get co-mentioned during the sample period. For all the linked pairs (i, j) identified throughout the sample period, 49.6% of them are only mentioned in one yearly window. We consider those pairs as temporarily linked. They could get co-mentioned multiple times within a yearly window. But out of that one-year window, they are never mentioned together. To further reduce noise, we say a pair (i, j) has persistent economic relationships if they are identified in more than a certain number $(1 \leqslant m \leqslant 11)$ of yearly identification windows. For the construction of full sample adjacency matrix W, we set w_{ij} to the number of times i and j are co-mentioned throughout the sample if the pair (i, j) gets co-mentioned in more than m yearly identification windows (i.e., their link is persistent), and to zero otherwise.

Table 12 presents the number of identified pairs aggregated at industry level for threshold $m = 1$. Results for higher threshold values are shown in Table 13, and Table 14. We classify stocks into Fama-French 12 industries based on their Standard Industrial Classification (SIC) code. Compared with companies from other industries, financial companies, hi-tech companies, and manufacturing companies are more connected. Another important feature is that there are a lot of intra-industry links. Except for some industries with very few stock like Durables and Telecommunication, whose statistics should be interpreted with care, other industries all have a high percentage of intra-industry links. Comparing tables of adjacency matrices with different threshold values m, we can tell that although higher threshold values reduce the absolute number of identified pairs, the relative industry level network remains very robust.

2.6 Results

2.6.1 Main Results

For full sample estimation, we keep *S&P* 500 stocks that have no missing observations for the period 2004 to 2015, which leaves us $N = 394$ stocks. Adjacency matrix W contains all the persistent links (for different thresholds m) identified throughout the sample. As a convention in spatial econometrics, we then apply row-normalization to W so that $\sum_j^N w_{ij} = 1$ for all $i = 1, \ldots, N$. We investigate several models under the general framework Equation 2.25:

- Model 1: Spatial CAPM model

$$\tilde{\boldsymbol{r}}_t = \tilde{\boldsymbol{\alpha}} + \boldsymbol{\beta_1} f_{MRT,t} + \sum_{l=0}^{L} \Psi_l W \tilde{\boldsymbol{r}}_{t-l} + \boldsymbol{\varepsilon}_t. \tag{2.27}$$

- Model 2: Spatial factor model with Fama-French three factors

$$\tilde{\boldsymbol{r}}_t = \tilde{\boldsymbol{\alpha}} + \boldsymbol{\beta_1} f_{MRT,t} + \boldsymbol{\beta_2} f_{SMB,t} + \boldsymbol{\beta_3} f_{HML,t} + \sum_{l=0}^{L} \Psi_l W \tilde{\boldsymbol{r}}_{t-l} + \boldsymbol{\varepsilon}_t. \tag{2.28}$$

- Model 3: Spatial factor model with Fama-French five factors

$$\tilde{\boldsymbol{r}}_t = \tilde{\boldsymbol{\alpha}} + \boldsymbol{\beta_1} f_{MRT,t} + \boldsymbol{\beta_2} f_{SMB,t} + \boldsymbol{\beta_3} f_{HML,t} + \boldsymbol{\beta_4} f_{RMW,t} + \boldsymbol{\beta_5} f_{CMA,t} + \sum_{l=0}^{L} \Psi_l W \tilde{\boldsymbol{r}}_{t-l} + \boldsymbol{\varepsilon}_t. \tag{2.29}$$

- Model 4: Spatial factor model with Fama-French five factors plus Momentum factor

$$\tilde{\boldsymbol{r}}_t = \tilde{\boldsymbol{\alpha}} + \boldsymbol{\beta}_1 f_{MRT,t} + \boldsymbol{\beta}_2 f_{SMB,t} + \boldsymbol{\beta}_3 f_{HML,t} + \boldsymbol{\beta}_4 f_{RMW,t} + \boldsymbol{\beta}_5 f_{CMA,t} + \boldsymbol{\beta}_6 f_{MOM,t} + \sum_{l=0}^{L} \Psi_l W \tilde{\boldsymbol{r}}_{t-l} + \boldsymbol{\varepsilon}_t. \tag{2.30}$$

The parameters $(N*(K+L+3))$ in Equation 2.25 are estimated using quasi maximum likelihood (QML). Given the huge amount of parameters in the model, here we only report some important summary statistics of the estimates in Table 2.1.[2] For a heterogeneous coefficient panel model, what is often of interest to empirical researchers is the average estimates across all entities (or all entities within a sub-group). If we assume that individual-specific coefficients are randomly distributed around their common means as follows:

$$\begin{aligned} &\beta_{k,i} = \lambda_k + \zeta_{k,i}, \psi_{l,i} = \psi_l + \varsigma_{l,i} \quad \text{for } k = 1,\ldots,K, \, l = 1,\ldots,L, \text{ and } i = 1,\ldots,N. \\ &\boldsymbol{\eta}_i = (\boldsymbol{\zeta}_i', \boldsymbol{\varsigma}_i'')' \sim IID(\mathbf{0}, \Omega_\eta). \end{aligned} \tag{2.31}$$

The common mean parameters β_k and ψ_l for $k = 1,\ldots,K$, $l = 1,\ldots,L$ are the the objects of interest and they can be consistently estimated with the following mean group (MG) estimator given N and T are large, with $\sqrt{N}/T \to 0$. [3].

$$\hat{\beta}_{k1}^{MG} = \frac{1}{N}\sum_{i=1}^{N} \hat{\beta}_{k1,i} \;\; and \;\; \hat{\psi}_l^{MG} = \frac{1}{N}\sum_{i=1}^{N} \hat{\psi}_{l,i}. \tag{2.32}$$

For a heterogeneous coefficient spatial model, we can only identify the spatial coefficients of those units with at least one link. Spatial coefficients of those units with zero link cannot be identified, and we need to restrict them to be zeros. If we apply the full sample adjacency matrix W discussed above with threshold value $m = 1$, only $N_0 = 7$ out of $N = 394$ companies do not have any long-run links. $N_p = N - N_0 = 387$ units have unrestricted spatial coefficients. In contrast, individual-specific factor coefficients and intercepts are identified for all units, with $N_p = N = 394$.

We estimate Model 1 - Model 4 over the full sample period. We report the mean group (MG) estimates with their standard errors and the percentages of companies with statistically

[2] Full estimation results can be requested from the author.

[3] See Pesaran and Smith (1995) for proofs of the consistency when individual-specific coefficients are independently distributed. Recent development by Chudik and Pesaran (2019) proves the consistency under weakly correlated individual-specific estimators. In both cases, T and N are required to be large with $\sqrt{N}/T \to 0$. Intuitively, big T is required for the consistent estimation of individual-specific coefficients, and N needs to be big enough for the consistent estimation of the means. To see how the MG estimators behave in the context of heterogeneous spatial-temporal model, see Aquaro et al. (2020)

significant parameters at 5% level for models estimated with threshold value $m = 1$ in Table 2.1. Results for alternative thresholds m are reported in Table 15 and Table 16.

	(1) factor component							(2) spatial-temporal component					
	α	β_1	β_2	β_3	β_4	β_5	β_6	ψ_0	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5
(1) Spatial CAPM													
MG	0.015	0.564						0.446	0.002	-0.008	0.001	-0.003	0.004
	(0.001)	(0.022)						(0.020)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
%sig	9.1%	90.4%						89.9%	51.9%	28.2%	20.7%	30.2%	21.7%
N_p	394	394						387	387	387	387	387	387
(2) Spatial factor model (Fama-French three factors)													
MG	0.013	0.529	0.129	-0.137				0.489	0.002	-0.008	-0.001	-0.001	0.003
	(0.001)	(0.021)	(0.014)	(0.022)				(0.019)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	6.1%	86.0%	75.1%	82.7%				89.7%	54.0%	27.1%	21.4%	29.7%	18.9%
N_p	394	394	394	394				387	387	387	387	387	387
(3) Spatial factor model (Fama-French five factors)													
MG	0.011	0.544	0.144	-0.137	0.140	0.179		0.493	0.007	-0.007	-0.001	-0.002	0.002
	(0.001)	(0.021)	(0.014)	(0.023)	(0.022)	(0.021)		(0.019)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	5.6%	86.8%	73.9%	82.5%	74.6%	73.6%		89.1%	52.2%	27.6%	22.2%	27.9%	18.6%
N_p	394	394	394	394	394	394		387	387	387	387	387	387
(4) Spatial factor model (Fama-French five factors plus Momentum)													
MG	0.015	0.471	0.124	-0.116	0.074	0.113	-0.010	0.545	0.002	-0.014	0.002	-0.008	0.005
	(0.001)	(0.022)	(0.014)	(0.021)	(0.022)	(0.021)	(0.007)	(0.019)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	4.8%	86.5%	73.3%	79.2%	75.1%	74.1%	59.4%	88.1%	52.7%	27.4%	23.3%	27.9%	18.6%
N_p	394	394	394	394	394	394	394	387	387	387	387	387	387

Table 2.1 **QML estimation results of Equation 2.27 to Equation 2.30 using full sample.** Note: threshold $m = 1$. For each panel , the first row gives the mean group (MG) estimates for each parameter with their standard errors in the parenthesis. The row of each panel gives the percentages of unrestricted units with statistically significant parameters at 5% level, and the last row gives the number of unrestricted units N_p for each parameter.

Contemporaneous local dependency parameter ψ_0 is highly statistically significant under all specifications. Among 387 unrestricted contemporaneous spatial coefficients $\psi_{0,i}$, more than 88% of them are individually significant under all cases. This high significance ratio implies that the new-implied linkage identification method is very successful at detecting relevant links. If our data contain a lot of spurious links, we will be more likely to see the spatial parameters to be insignificant for many individuals. Local dependencies also exhibit strong economic importance: the mean group (MG) estimates of ψ_0 are around $0.45 - 0.55$ over the four models we consider, which is comparable to the average strength of the market factor, with the mean group (MG) estimates of market beta lying between $0.47 - 0.56$ across models.

Dynamic spatial dependency terms are also statistically significant, although smaller in economic magnitude. For the first lag ψ_1, there are more than 50% of $\psi_{1,i}$ are individually

significant under all cases. For further lags, there are always more than 20% of $\psi_{l,i}$ $(l = 2,\ldots,5)$ are individually significant.

Adding more common risk factors does not weaken local dependencies. The magnitudes of mean group estimates and the percentages of companies with statistically significant parameters at 5% level do not change with the number of factors we include. Interestingly, the magnitude of average contemporaneous local dependency captured by ψ_0^{MG} is the largest for Model 4, while the magnitude of average exposures to market factor is the smallest.

A large proportion of the news-implied links that we identify are intra-industry links. It has been documented widely that stocks within the same industry exhibit excess co-movement beyond common risk factors at market level (Moskowitz and Grinblatt (1999), Fan et al. (2016a), Engelberg et al. (2018)). In order to control for industry factors as an additional source of co-movement, we further augment Equation 2.27 to Equation 2.30 with industry factors.

$$\tilde{\boldsymbol{r}}_t = \tilde{\boldsymbol{\alpha}} + \mathbf{B}\tilde{\mathbf{f}}_\mathbf{t} + \boldsymbol{\beta}_I \tilde{f}_{IND,t} + \sum_{l=0}^{L} \Psi_l W \tilde{\boldsymbol{r}}_{t-l} + \boldsymbol{\varepsilon}_t. \tag{2.33}$$

We use Fama French 12 equal-weighted industry portfolios. We choose to use broad industry classification and equal weighting. This is because we are dealing with *S&P* 500 stocks, and we do not want industry returns to be dominated by several large stocks within that industry.

Table 2.2 reports the results for models with industry factors. Industry factors are highly significant in all cases, and the mean group (MG) estimates of industry beta are between $0.41 - 0.45$. The introduction of the industry factor largely weakens the effect of the market factor, with the average market beta being reduced to $0.20 - 0.23$. On the other hand, the magnitudes of local dependencies are only slightly reduced by the introduction of the industry factor. This shows that our results are not driven by exposure to common industry-level shocks but by granular interactions. Using other equal-weighted industry factors does not affect this finding.

	(1) factor component							(2) spatial-temporal component						
	α	β_1	β_2	β_3	β_4	β_5	β_6	β_I	ψ_0	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5
(1) Spatial CAPM+ Industry factor														
MG	0.009	0.231						0.407	0.387	-0.015	-0.024	-0.008	-0.008	0.002
	(0.002)	(0.025)						(0.021)	(0.018)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
%sig	7.9%	81.0%						84.5%	87.3%	49.6%	46.0%	27.9%	30.5%	21.7%
N_P	394	394						394	387	387	387	387	387	387
(2) Spatial factor model (Fama-French three factors+ Industry factor)														
MG	0.007	0.197	-0.120	-0.156				0.451	0.415	-0.018	-0.019	-0.009	-0.007	-0.000
	(0.002)	(0.026)	(0.018)	(0.018)				(0.026)	(0.018)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
%sig	8.1%	79.2%	72.8%	82.0%				81.7%	86.6%	51.7%	39.8%	27.1%	28.4%	19.6%
N_p	394	394	394	394				394	387	387	387	387	387	387
(3) Spatial factor model (Fama-French five factors+ Industry factor)														
MG	0.005	0.212	-0.108	-0.164	0.106	0.199		0.445	0.422	-0.014	-0.018	-0.009	-0.008	-0.001
	(0.002)	(0.025)	(0.018)	(0.019)	(0.019)	(0.017)		(0.025)	(0.018)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
%sig	6.9%	77.7%	71.6%	81.5%	69.0%	70.8%		82.0%	86.6%	46.8%	36.7%	28.2%	28.4%	20.4%
N_p	394	394	394	394	394	394		394	387	387	387	387	387	387
(4) Spatial factor model (Fama-French five factors plus Momentum+ Industry factor)														
MG	0.005	0.221	-0.104	-0.155	0.102	0.194	0.003	0.435	0.420	-0.013	-0.017	-0.009	-0.008	-0.001
	(0.002)	(0.026)	(0.018)	(0.018)	(0.019)	(0.017)	(0.007)	(0.025)	(0.018)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
%sig	6.3%	79.4%	71.1%	80.2%	69.0%	67.8%	54.1%	80.7%	85.5%	46.0%	37.2%	27.6%	27.9%	20.4%
N_p	394	394	394	394	394	394	394	394	387	387	387	387	387	387

Table 2.2 **QML estimation results of industry factors augmented models (Equation 2.33) using full sample.**
Note: W is constructed using threshold $m = 1$. Estimation results for alternative threshold values are reported in Table 17 and Table 18.

So far, we have shown the mean group (MG) estimation results for whole sample companies. It is also interesting to gauge heterogeneity at sub-group levels. It is reasonable to suspect that the mean sensitivities to local risk spillovers are different for different industry groups. To explore this heterogeneity, here we adopt the random coefficient assumptions at the industry level. Subscript g denotes industry membership, and we classify stocks into six broad industries[4].

$$\begin{aligned} &\beta_{k,i,g} = \beta_{k,g} + \zeta_{k,i,g}, \quad \psi_{l,i,g} = \psi_{l,g} + \varsigma_{l,i,g} \\ &\text{for } k = 1,\ldots,K,\ l = 1,\ldots,L \text{ ,and } i = 1,\ldots,N,\ g = 1,\ldots,G. \\ &\boldsymbol{\eta}_{i,g} = (\boldsymbol{\zeta}'_{i,g}, \boldsymbol{\varsigma}'_{i,g}{}')' \sim IID(\mathbf{0}, \Omega_\eta). \end{aligned} \tag{2.34}$$

[4]We adopt broad industry classification to guarantee that there are a large number of stocks within each industry since mean group estimation requires large N to be consistent. We build the industry classification on top of the Fama-French five industry definitions where they classify all stocks according to their SIC code into five broad groups: "Consumer", "Health", "Hi-tech", "Manufacturing" and "Others". For the first four categories, we keep the same definitions as Fama and French. Since there are a large proportion of financial companies in the *S&P*500 universe, it would be interesting to separate financial firms from those in the "Others" category. Among the stocks that fall into "Others", we categorize the stocks with a SIC in the range $6000-6799$ as "Finance" and put the remaining in the "Others" category.

The industry-level common mean parameters for industry g can be consistently estimated when N_g, the number of cross-sectional units within that industry is large.

$$\hat{\beta}_{k,g}^{MG} = \frac{1}{N_g} \sum_{i \in \mathbb{N}_g} \hat{\beta}_{k,i,g} \quad and \quad \hat{\psi}_{l,g}^{MG} = \frac{1}{N_g} \sum_{i \in \mathbb{N}_g} \hat{\psi}_{l,i,g}. \tag{2.35}$$

We report the mean group (MG) estimates by industry for the spatial factor model with Fama-French five factors plus the momentum factor Equation 2.30 and its counterpart with the industry factor in Table 2.3 and Table 2.4, respectively. Both tables reveal that our main conclusions that equity returns are affected by that of their economic neighbours are very robust to the industry disaggregation. Local dependencies are highly significant for all six industries. The industrial mean group (MG) estimates of ψ_0 are between $0.36 - 0.58$ for the Spatial factor model with Fama-French five factors and the momentum factor (Equation 2.30). After controlling for the industry factor, the estimates still range from $0.31 - 0.56$.

Financial companies have the largest exposures to their neighbours' shocks. And this high level of sensitivity to local shocks cannot be explained by exposures to common industry shocks as the estimates of spatial parameters stay unchanged with the introduction of the industry factor. After controlling for the industry factor, the mean group (MG) estimates of ψ_0 for the financial industry is still as large as $0.56\ (0.05)$. By contrast, the introduction of the industry factor reduces the estimated local dependencies for the consumer industry, health industry, and manufacturing industry by a larger margin. Apart from the large contemporaneous spatial coefficient, it is also worth noticing that financial companies also have a stronger momentum spillover effect. The percentage of significant spatial-temporal coefficients[5] for financial companies are much larger than that for other industries at any lag.

[5] We need to interpret the mean group estimates of these spatial-temporal parameters with care. The individual parameters $\psi_{l,i,g}$ are quite dispersed for $l \geqslant 1$, with some firms having significantly positive spatial-temporal terms and some having significantly negative ones. That is why the mean group estimates for these spatial-temporal parameters may not look very statistically significant, although high percentages of individual coefficients are significant — there is simply too much heterogeneity.

	(1) factor component							(2) spatial-temporal component					
	α	β_1	β_2	β_3	β_4	β_5	β_6	ψ_0	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5
Panel A: Finance													
MG	0.017	0.471	0.073	0.410	-0.002	-0.086	-0.101	0.584	-0.040	-0.013	-0.008	-0.007	0.003
	(0.002)	(0.055)	(0.037)	(0.068)	(0.058)	(0.050)	(0.019)	(0.044)	(0.010)	(0.005)	(0.004)	(0.005)	(0.004)
%sig	5.9%	73.5%	67.6%	86.8%	55.9%	54.4%	64.7%	88.2%	75.0%	47.1%	36.8%	47.1%	29.4%
N_p	68	68	68	68	68	68	68	68	68	68	68	68	68
Panel B: Consumer													
MG	0.012	0.525	0.186	-0.230	0.303	0.401	-0.048	0.412	0.019	-0.012	-0.002	0.001	-0.000
	(0.003)	(0.044)	(0.033)	(0.023)	(0.027)	(0.030)	(0.015)	(0.040)	(0.005)	(0.004)	(0.004)	(0.004)	(0.003)
%sig	5.3%	85.3%	80.0%	76.0%	82.7%	89.3%	53.3%	86.7%	48.0%	26.7%	28.0%	28.0%	17.3%
N_p	75	75	75	75	75	75	75	75	75	75	75	75	75
Panel C: Health													
MG	0.026	0.495	0.025	-0.396	-0.149	0.161	0.085	0.364	0.016	-0.003	0.013	0.004	0.005
	(0.007)	(0.066)	(0.044)	(0.038)	(0.053)	(0.046)	(0.022)	(0.058)	(0.007)	(0.004)	(0.004)	(0.005)	(0.005)
%sig	21.4%	92.9%	64.3%	96.4%	64.3%	60.7%	67.9%	89.3%	39.3%	7.1%	17.9%	10.7%	10.7%
N_p	28	28	28	28	28	28	28	28	28	28	28	28	28
Panel D: Hi-tech													
MG	0.016	0.651	0.163	-0.410	-0.292	0.280	-0.038	0.407	0.009	-0.009	0.005	-0.002	0.007
	(0.004)	(0.045)	(0.032)	(0.028)	(0.047)	(0.042)	(0.013)	(0.043)	(0.006)	(0.004)	(0.004)	(0.004)	(0.003)
%sig	5.7%	91.4%	65.7%	92.9%	68.6%	65.7%	45.7%	85.5%	55.1%	27.5%	17.4%	18.8%	14.5%
N_p	70	70	70	70	70	70	70	69	69	69	69	69	69
Panel E: Manufacturing													
MG	0.002	0.538	0.129	-0.169	0.424	0.121	0.049	0.580	0.019	-0.000	-0.002	-0.001	-0.002
	(0.002)	(0.045)	(0.024)	(0.022)	(0.026)	(0.047)	(0.013)	(0.038)	(0.004)	(0.002)	(0.002)	(0.003)	(0.002)
%sig	0.0%	89.4%	80.5%	66.4%	93.8%	81.4%	69.9%	91.6%	49.5%	20.6%	14.0%	27.1%	17.8%
N_p	113	113	113	113	113	113	113	107	107	107	107	107	107
Panel F: Others													
MG	0.007	0.634	0.288	-0.217	0.206	0.258	-0.050	0.418	0.013	-0.007	-0.010	-0.010	-0.000
	(0.005)	(0.070)	(0.043)	(0.052)	(0.046)	(0.053)	(0.021)	(0.067)	(0.006)	(0.004)	(0.004)	(0.005)	(0.004)
%sig	2.5%	90.0%	72.5%	72.5%	60.0%	80.0%	50.0%	85.0%	37.5%	27.5%	30.0%	25.0%	17.5%
N_p	40	40	40	40	40	40	40	40	40	40	40	40	40

Table 2.3 **QML estimation results of Spatial factor model with Fama-French five factors and the momentum factor (Equation 2.30). Parameters summarized by industry.**

	(1) factor component								(2) spatial-temporal component					
	α	β_1	β_2	β_3	β_4	β_5	β_6	β_I	ψ_0	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5
Panel A: Finance														
MG	0.010	0.298	-0.068	0.313	0.036	-0.055	-0.071	0.302	0.560	-0.041	-0.018	-0.013	-0.009	0.001
	(0.003)	(0.063)	(0.058)	(0.054)	(0.055)	(0.048)	(0.018)	(0.062)	(0.046)	(0.010)	(0.005)	(0.005)	(0.005)	(0.004)
%sig	7.4%	89.7%	72.1%	88.2%	58.8%	54.4%	67.6%	66.2%	85.3%	75.0%	50.0%	39.7%	42.6%	33.8%
N_p	68	68	68	68	68	68	68	68	68	68	68	68	68	68
Panel B: Consumer														
MG	0.003	0.084	-0.195	-0.226	0.235	0.318	0.015	0.576	0.358	-0.010	-0.025	-0.014	-0.008	-0.008
	(0.003)	(0.056)	(0.028)	(0.022)	(0.024)	(0.028)	(0.012)	(0.053)	(0.037)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
%sig	4.0%	74.7%	70.7%	77.3%	74.7%	76.0%	52.0%	90.7%	81.3%	40.0%	38.7%	30.7%	28.0%	16.0%
N_p	75	75	75	75	75	75	75	75	75	75	75	75	75	75
Panel C: Health														
MG	0.011	0.280	-0.200	-0.262	0.116	0.228	0.071	0.361	0.305	-0.014	-0.022	0.000	-0.005	-0.004
	(0.007)	(0.054)	(0.039)	(0.039)	(0.040)	(0.044)	(0.022)	(0.036)	(0.059)	(0.007)	(0.004)	(0.004)	(0.005)	(0.005)
%sig	10.7%	78.6%	67.9%	96.4%	42.9%	67.9%	64.3%	96.4%	92.9%	28.6%	14.3%	10.7%	7.1%	14.3%
N_p	28	28	28	28	28	28	28	28	28	28	28	28	28	28
Panel D: Hi-tech														
MG	0.017	0.628	0.147	-0.396	-0.301	0.261	-0.035	-0.002	0.410	0.006	-0.009	0.004	-0.004	0.007
	(0.004)	(0.043)	(0.040)	(0.029)	(0.049)	(0.043)	(0.013)	(0.034)	(0.041)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
%sig	7.1%	85.7%	64.3%	90.0%	72.9%	62.9%	40.0%	58.6%	84.1%	42.0%	27.5%	15.9%	18.8%	10.1%
N_p	70	70	70	70	70	70	70	70	69	69	69	69	69	69
Panel E: Manufacturing														
MG	-0.002	-0.028	-0.246	-0.188	0.249	0.189	0.053	0.732	0.407	-0.014	-0.016	-0.012	-0.009	-0.002
	(0.002)	(0.048)	(0.029)	(0.020)	(0.020)	(0.030)	(0.011)	(0.047)	(0.036)	(0.004)	(0.003)	(0.002)	(0.003)	(0.002)
%sig	6.2%	75.2%	81.4%	69.0%	76.1%	69.9%	54.9%	94.7%	87.9%	47.7%	40.2%	27.1%	29.0%	24.3%
N_p	113	113	113	113	113	113	113	113	107	107	107	107	107	107
Panel F: Others														
MG	-0.000	0.299	0.031	-0.225	0.248	0.257	-0.019	0.376	0.432	-0.006	-0.017	-0.016	-0.014	-0.003
	(0.005)	(0.086)	(0.044)	(0.050)	(0.047)	(0.053)	(0.021)	(0.054)	(0.062)	(0.006)	(0.004)	(0.004)	(0.005)	(0.004)
%sig	5.0%	72.5%	55.0%	75.0%	67.5%	77.5%	50.0%	75.0%	85.0%	22.5%	37.5%	35.0%	30.0%	17.5%
N_p	40	40	40	40	40	40	40	40	40	40	40	40	40	40

Table 2.4 **QML estimation results of spatial factor model with Fama-French five factors, the momentum factor, and the industry factor. Parameters summarized by industry.**

Next, we examine how the spatial factor model captures the remaining dependence in the de-factored returns. Using the method described in subsection 2.2.3, we compute the number of non-zero pair-wise cross correlations of residuals from (1) Factor Model with FF5+MOM+IF that use Fama-French five factors, and momentum factor + 12 Industry factor, and (2) Spatial Factor Model with FF5+MOM+IF that use Fama-French five factors, and momentum factor + 12 Industry factor. [6]

To test $H_{0,ij} : \rho_{ij} = 0$ for $n_{test} = N(N-1)/2$ pairs of (i, j), for a given family-wise error rate (FWER) p, the critical values are $\pm\frac{1}{\sqrt{T}}\Phi^{-1}(1 - p/2n_{test})$ for the factor model, and $F^{-1}(p/2n_{test})$ and $F^{-1}(1 - p/2n_{test})$ for the spatial factor model. F is the empirical null distribution from $B = 500$ bootstrap samples. Figure 1 shows the histogram of bootstrapped $\hat{\rho}_{ij}^b$ for all $i \neq j, b = 1, \ldots, 500$ for the spatial factor model. Table 2.5 presents the degree of cross-sectional dependence in the factor model and its spatial-augmented version under different family-wise error rates. First of all, bootstrap results show that the limiting distribution of $\hat{\rho}_{ij}$ under the null is indeed altered by the addition of the spatial term, and we

[6] Here we only present the results for the models with most factors that are supposed to have least residual correlations among all. We can do more different factor models and their spatial augmented versions, at the cost of bootstrap inference for each spatial factor model.

need different critical values for testing. The table shows that adding the spatial component reduces the number of non-zero pair-wise cross correlations by a huge margin.[7] The spatial component constructed with news-implied linkages is successful at eliminating correlations from the de-factored returns.

	Critical values	# Non-zero pair-wise cross correlations	Density
(1) $p = 0.05$			
Factor Model with FF5+MOM+IF	-0.091,0.091	8516	5.50%
Spatial Factor Model with FF5+MOM+IF	-0.243,0.245	416	0.27%
(2) $p = 0.1$			
Factor Model with FF5+MOM+IF	-0.088,0.088	9490	5.87%
Spatial Factor Model with FF5+MOM+IF	-0.225,0.229	478	0.31%

Table 2.5 Degree of Cross-Sectional Dependence in the Residuals.
Note: Density gives the percentage of non-zero pair-wise cross correlations (i.e., density=number of non-zero pair-wise cross-correlations/N(N-1)).

2.6.2 Model Performance: Degree of Mispricing

Next, we compare the performance of factor models with their spatial versions by several measures of the degree of mispricing. We are interested in whether adding the spatial component to the factor model could reduce mispricing.

If an asset pricing model completely captures expected returns, then the intercept $\tilde{\boldsymbol{\alpha}}$ is approximately zero. We are dealing with a high-dimensional system. Testing $H_0 : \tilde{\boldsymbol{\alpha}} = 0$ for large N using traditional tests like Gibbons et al. (1989) which are designed for cases where the number of test assets are small will have low power problem. Also, this test for whether exact arbitrage pricing holds is stronger than what is implied by the approximate no arbitrage condition (Equation 2.20 and Equation 2.21).

So instead of using standard GRS test, we employ other three statistics to compare the relative performance of different models (1) the percentage of individually significant $\tilde{\alpha}_i$ (Pesaran and Yamagata (2012)); (2) average L_1 norm of intercepts $A(\mid \tilde{\alpha}_i \mid)$; (3) average L_2 norm of intercepts $A(\tilde{\alpha}_i^2)$ (Fama and French (2015)). Those three statistics are implied by 1 and the corollaries to be useful in measuring how well the approximation (Equation 2.20) is. In addition to those three measures of mispricing, we also report the mean-squared errors (MSE) of different models.

[7] This is not only because we need larger critical values under the spatial factor specification. Even if we do not consider the distortion brought by the spatial component and still use limiting distributions from factor model residual correlation coefficients, the percentage of non-zero pair-wise cross correlations is still reduced by half.

We compare the relative performances of the following factor models, their spatial augmented versions.

- Model 1.1: CAPM model (CAPM)
- Model 1.2: CAPM model + Industry factor (CAPM+IF)
- Model 1.3: Spatial CAPM model (CAPM(S))
- Model 1.4: Spatial CAPM model + Industry factor (CPAM+IF(S))
- Model 2.1: Factor model with Fama-French three factors (FF3)
- Model 2.2: Factor model with Fama-French three factors + Industry factor (FF3+IF)
- Model 2.3: Spatial factor model with Fama-French three factors (FF3(S))
- Model 2.4: Spatial factor model with Fama-French three factors + Industry factor (FF3+IF(S))
- Model 3.1: Factor model with Fama-French five factors (FF5)
- Model 3.2: Factor model with Fama-French five factors + Industry factor (FF5+IF)
- Model 3.3: Spatial factor model with Fama-French five factors (FF5(S))
- Model 3.4: Spatial factor model with Fama-French five factors + Industry factor (FF5+IF(S))
- Model 4.1: Factor model with Fama-French five factors, and momentum factor (FF5+MOM)
- Model 4.2: Factor model with Fama-French five factors, and momentum factor + Industry factor (FF5+MOM+IF)
- Model 4.3: Spatial factor model with Fama-French five factors, and momentum factor (FF5+MOM(S))
- Model 4.4: Spatial factor model with Fama-French five factors, and momentum factor + Industry factor (FF5+MOM+IF(S))

	% of significant $\tilde{\alpha}_i$	$A(\mid \tilde{\alpha}_i \mid)$	$A(\tilde{\alpha}_i^2)$	Mean-squared error (MSE)
Model 1.1: CAPM	7.61%	2.51	10.91	3.26
Model 1.2: CAPM+IF	7.61%	2.47	11.19	2.99
Model 1.3: CAPM(S)	9.13%	2.43	10.41	2.86
Model 1.4: CAPM+IF(S)	7.87%	2.36	10.41	2.73
Model 2.1: FF3	7.64%	2.49	10.66	3.26
Model 2.2: FF3+IF	7.35%	2.35	10.69	2.86
Model 2.3: FF3(S)	6.09%	2.31	9.39	2.75
Model 2.4: FF3+IF(S)	8.12%	2.28	9.90	2.65
Model 3.1: FF5	5.84%	2.45	11.17	2.99
Model 3.2: FF5+IF	6.60%	2.39	11.42	2.82
Model 3.3: FF5(S)	5.58%	2.21	9.39	2.72
Model 3.4: FF5+IF(S)	6.68%	2.24	9.90	2.62
Model 4.1: FF5+MOM	6.09%	2.49	11.18	2.97
Model 4.2: FF5+MOM+IF	6.35%	2.39	11.34	2.80
Model 4.3: FF5+MOM(S)	4.80%	2.18	9.14	2.70
Model 4.4: FF5+MOM+IF(S)	6.45%	2.19	9.64	2.61

Table 2.6 Summary of Model Performance.
Note: Each panel shows the performance statistics of a factor model, its spatial augmented version, and its spatial and industry factor augmented version. Note: $\tilde{\boldsymbol{\alpha}}$ used to compute $A(\mid \tilde{\alpha}_i \mid)$ and $A(\tilde{\alpha}_i^2)$ are in basis point, and $\boldsymbol{\varepsilon}$ used to compute MSE errors are all in percentage point. For each column, the best statistic is highlighted in red.

Table 2.6 shows that for all factor models except the CAPM, adding spatial interactions improves all performance measures. Spatial CAPM fails to reduce the percentage of individually significant intercepts. This is because spatial models are designed for modelling local interactions. If there are not enough common risk factors to capture the strong dependence in equity returns, adding the spatial component which deals with weak dependence is not going to be helpful. For factor models with Fama-French three-factor, five-factor, and five-factor plus momentum factor, adding spatial interactions all provides noticeable improvement on reducing the mispricing component and mean squared errors. Interestingly, although adding the industry factor can further reduce mean-squared error, it cannot help to reduce the mispricing [8].

[8] In unreported tables, if we replace equal-weighted industry portfolios with value-weighted industry portfolios, the introduction of industry factor does further bring down three statistics of mispricing. And model 4.4, the spatial factor model with Fama-French five factors, and momentum factor plus Industry factor has the

For each measure, the best-performing statistic is highlighted in red. Model 4.3, the spatial factor model with Fama-French five factors, and momentum factor appears to have the best performance in terms of reducing pricing errors. And model 4.4, which is model 4.3 with the industry factor, has the smallest mean-squared errors.

2.6.3 Comparisons with alternative networks

In this section, we gauge whether the news-implied links carry additional information on top of existing linkage datasets. We first show that spatial factor models estimated with W constructed using other existing linkage datasets under-perform that estimated with news-implied W. We then show that conditional on other existing linkages, local risk spillovers via our news-implied links continue to be significant.

We consider the following competing networks:

- Industry-based adjacency matrices based on industry classification of different granularities including 4-digit SIC codes, 3-digit SIC codes, and 2-digit SIC codes classifications. This is motivated by Moskowitz and Grinblatt (1999), Engelberg et al. (2018) and Fan et al. (2016a). For each classification criteria, we build block-diagonal matrices where companies within the same industry are fully connected.

- IBES analyst co-coverage networks. It has been documented that shared analyst coverage is a strong proxy for fundamental linkages between firms and reflects firm similarities along many dimensions (Ali and Hirshleifer (2020), Israelsen (2016), Kaustia and Rantala (2013)). We use the Institutional Brokers Estimate System (IBES) detail history files to construct the analyst co-coverage-based adjacency matrix. For each year in the sample, we consider a stock is covered by an analyst if the analyst issues at least one FY1 or FY2 earnings forecast for the stock during the year. And we consider two stocks as linked if there are common analysts during the year, weighted by the number of common analysts. We then add up the yearly adjacency matrices to get the full sample adjacency matrix.

- Customer-supplier links (Cohen and Frazzini (2008)) from Andrea Frazzini's data library. The strength of links is weighted by sales.

best performance in all dimensions. However, as we have argued earlier, value-weighted portfolios might cause endogeneity issues given we are working with large companies. Using other equal weighting industry factors does not change the results.

- Geographic links (Pirinsky and Wang (2006) and Parsons et al. (2020)). We obtain location information from CRSP Compustat merged files. We then merge the sample firms with the Metropolitan Statistical Areas (MSA) data using the ZIP-FIPS-MSA data from the US Department of Labor, which maps zip codes to MSAs. We follow Pirinsky and Wang (2006), and consider firms whose headquarters are in the same MSA as linked.

- The union of above mentioned links. We let a typical entry w_{ij} in this matrix to be one if the pair (i, j) is linked in any of the above networks, and zero otherwise.

Table 2.7 shows the performance of competing networks. Individually, spatial APT models estimated with new-implied linkages out-perform other existing networks. And even if we consider the union of all alternative networks, the news-implied network still does not lose the battle. For the four statistics that we consider, spatial factor model with Fama-French five factors and momentum factor estimated using W_{news} (last line of panel (1)) performs the best among all candidates along two dimensions. Spatial factor model with Fama-French five factors, momentum factor, and industry factor estimated using W_{union} (second last line of panel (2)) also performs the best among all candidates along two dimensions. Models with adjacency matrices capturing multiple channels out-perform those with adjacency matrices focusing on one particular channel. This seems to support the fact that there are multiple channels of local risk spillovers. There is no reason to focus on one particular channel like intra-industry channel, customer-supplier channel, etc. Our news-implied linkages provide a comprehensive and integrated measure of firm-level relatedness, and it can be seen as a nice proxy of firm-to-firm connectivity.

	% of significant $\tilde{\alpha}_i$	$A(\mid \tilde{\alpha}_i \mid)$	$A(\tilde{\alpha}_i^2)$	Mean-squared error (MSE)
Panel (1): Spatial factor model with Fama-French five factors, and momentum factor				
$W_{2-digit-SIC}$	7.61%	2.22	9.64	2.68
$W_{3-digit-SIC}$	6.85%	2.37	10.91	2.82
$W_{4-digit-SIC}$	6.35%	2.40	11.68	2.88
W_{IBES}	6.85%	2.29	9.90	2.73
$W_{Customer-Supplier}$	6.09%	2.48	11.17	2.96
$W_{Geographic}$	7.11%	2.67	12.44	2.97
W_{Union}	6.09%	2.17	9.14	2.64
W_{News}	4.80%	2.18	9.14	2.70
Panel (2): Spatial factor model with Fama-French five factors, and momentum factor + Industry factor				
$W_{2-digit-SIC}$	6.09%	2.23	10.14	2.65
$W_{3-digit-SIC}$	5.58%	2.31	10.91	2.77
$W_{4-digit-SIC}$	5.58%	2.35	11.93	2.83
W_{IBES}	6.09%	2.22	9.90	2.68
$W_{Customer-Supplier}$	6.85%	2.39	11.17	2.81
$W_{Geographic}$	6.60%	2.45	11.93	2.81
W_{Union}	5.83%	2.16	9.39	2.61
W_{News}	6.45%	2.19	9.64	2.61

Table 2.7 Summary of Model Performance using competing networks.
Note: Panel (1) shows the performance of competing adjacency matrices under the spatial factor model with Fama-French five factors, and momentum factor. Panel (2) shows the performance of competing adjacency matrices under the spatial factor model with Fama-French five factors, and momentum factor plus industry factor. $\tilde{\boldsymbol{\alpha}}$ used to compute $A(\mid \tilde{\alpha}_i \mid)$ and $A(\tilde{\alpha}_i^2)$ are in basis point, and $\boldsymbol{\varepsilon}$ used to compute MSE errors are all in percentage point. For each column, the best statistic is highlighted in red.

Next, we examine that whether our news-implied linkages carry new information on top of existing linkages documented? To do that, we estimate the two-W spatial factor models below, with W_1 being our news-implied networks and W_2 being a set of other candidate matrices

$$\tilde{\boldsymbol{r}}_t = \tilde{\boldsymbol{\alpha}} + \mathbf{B}\tilde{\mathbf{f}}_\mathbf{t} + \sum_{l=0}^{L} \Psi_{1,l} W_1 \tilde{\boldsymbol{r}}_{t-l} + \sum_{l=0}^{L} \Psi_{2,l} W_2 \tilde{\boldsymbol{r}}_{t-l} + \boldsymbol{\varepsilon}_t. \tag{2.36}$$

$$\tilde{\boldsymbol{r}}_t = \tilde{\boldsymbol{\alpha}} + \mathbf{B}\tilde{\mathbf{f}}_\mathbf{t} + \boldsymbol{\beta}_I f_{IND,t} + \sum_{l=0}^{L} \Psi_{1,l} W_1 \tilde{\boldsymbol{r}}_{t-l} + \sum_{l=0}^{L} \Psi_{2,l} W_2 \tilde{\boldsymbol{r}}_{t-l} + \boldsymbol{\varepsilon}_t. \tag{2.37}$$

	(1) W_1						(2) W_2					
	$\psi_{1,0}$	$\psi_{1,1}$	$\psi_{1,2}$	$\psi_{1,3}$	$\psi_{1,4}$	$\psi_{1,5}$	$\psi_{2,0}$	$\psi_{2,1}$	$\psi_{2,2}$	$\psi_{2,3}$	$\psi_{2,4}$	$\psi_{2,5}$
Panel(1): $W_1 = W_{news}$ **and** $W_2 = W_{2-digit-SIC}$												
MG	0.299	0.005	-0.002	0.001	0.001	0.000	0.309	-0.000	-0.005	-0.004	-0.002	0.001
	(0.017)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.020)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
%sig	75.2%	32.3%	19.6%	21.2%	20.2%	19.6%	82.8%	27.1%	16.4%	18.3%	18.8%	17.8%
N_p	387	387	387	387	387	387	377	377	377	377	377	377
Panel(2): $W_1 = W_{news}$ **and** $W_2 = W_{3-digit-SIC}$												
MG	0.296	0.007	-0.003	-0.000	-0.002	-0.001	0.271	0.000	-0.003	-0.003	0.001	0.002
	(0.019)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.026)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
%sig	75.5%	35.9%	19.9%	20.9%	25.3%	17.8%	83.4%	26.8%	19.0%	17.5%	17.8%	15.1%
N_p	387	387	387	387	387	387	332	332	332	332	332	332
Panel(3): $W_1 = W_{news}$ **and** $W_2 = W_{4-digit-SIC}$												
MG	0.318	0.010	-0.003	-0.000	-0.003	0.000	0.264	-0.002	-0.004	-0.002	0.002	0.002
	(0.019)	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)	(0.029)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
%sig	75.2%	39.3%	19.9%	20.7%	24.3%	19.6%	81.6%	26.7%	18.8%	17.7%	17.7%	16.3%
N_p	387	387	387	387	387	387	288	288	288	288	288	288
Panel(4): $W_1 = W_{news}$ **and** $W_2 = W_{IBES}$												
MG	0.365	0.004	-0.005	-0.001	-0.003	-0.001	0.197	0.001	-0.001	-0.002	0.000	0.003
	(0.019)	(0.004)	(0.002)	(0.003)	(0.003)	(0.002)	(0.014)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
%sig	83.5%	39.8%	20.9%	19.1%	22.5%	19.4%	80.0%	26.2%	16.8%	17.6%	15.6%	17.6%
N_p	387	387	387	387	387	387	340	340	340	340	340	340
Panel(5): $W_1 = W_{news}$ **and** $W_2 = W_{Customer-Supplier}$												
MG	0.483	0.006	-0.007	-0.001	-0.003	0.001	0.082	0.005	-0.004	-0.003	0.008	0.003
	(0.019)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.021)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)
%sig	87.1%	50.6%	26.1%	23.3%	27.1%	17.6%	69.5%	13.6%	11.9%	11.9%	10.2%	13.6%
N_p	387	387	387	387	387	387	59	59	59	59	59	59
Panel(6): $W_1 = W_{news}$ **and** $W_2 = W_{Geographic}$												
MG	0.459	0.006	-0.006	0.002	-0.000	0.001	0.093	-0.002	-0.001	-0.005	-0.005	0.001
	(0.018)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.019)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
%sig	86.3%	36.7%	21.7%	19.9%	21.2%	15.8%	59.9%	22.5%	18.2%	12.1%	14.0%	13.4%
N_p	387	387	387	387	387	387	307	307	307	307	307	307
Panel(7): $W_1 = W_{news}$ **and** $W_2 = W_{Union}$												
MG	0.305	0.007	-0.001	0.003	0.003	-0.000	0.374	-0.004	-0.005	-0.005	-0.005	0.001
	(0.016)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.020)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
%sig	75.5%	26.6%	17.8%	20.2%	18.9%	16.3%	83.8%	23.6%	17.3%	18.3%	17.0%	17.5%
N_p	387	387	387	387	387	387	394	394	394	394	394	394

Table 2.8 **QML estimation results of two-W spatial factor model with Fama-French five factors, and the Momentum factor (Equation 2.36).**

Note: we only report spatial parameters here. W_{news} is constructed using threshold $m = 1$.

	(1) W_1						(2) W_2					
	$\psi_{1,0}$	$\psi_{1,1}$	$\psi_{1,2}$	$\psi_{1,3}$	$\psi_{1,4}$	$\psi_{1,5}$	$\psi_{2,0}$	$\psi_{2,1}$	$\psi_{2,2}$	$\psi_{2,3}$	$\psi_{2,4}$	$\psi_{2,5}$
Panel(1): $W_1 = W_{news}$ **and** $W_2 = W_{2-digit-SIC}$												
MG	0.280	-0.002	-0.007	-0.003	-0.003	-0.001	0.240	-0.011	-0.009	-0.007	-0.004	-0.001
	(0.017)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.019)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
%sig	74.7%	26.9%	21.2%	22.7%	21.2%	21.2%	82.0%	28.9%	15.4%	18.6%	17.8%	18.3%
N_p	387	387	387	387	387	387	377	377	377	377	377	377
Panel(2): $W_1 = W_{news}$ **and** $W_2 = W_{3-digit-SIC}$												
MG	0.267	-0.004	-0.011	-0.006	-0.006	-0.003	0.228	-0.007	-0.005	-0.004	-0.001	0.001
	(0.018)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.025)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
%sig	72.4%	31.3%	24.0%	24.8%	22.7%	18.1%	80.4%	28.3%	19.3%	18.1%	16.0%	16.0%
N_p	387	387	387	387	387	387	332	332	332	332	332	332
Panel(3): $W_1 = W_{news}$ **and** $W_2 = W_{4-digit-SIC}$												
MG	0.282	-0.003	-0.010	-0.007	-0.007	-0.002	0.222	-0.010	-0.006	-0.003	0.000	0.001
	(0.018)	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)	(0.028)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
%sig	71.3%	31.8%	25.3%	24.3%	22.0%	19.4%	78.8%	29.5%	20.1%	18.1%	18.1%	17.0%
N_p	387	387	387	387	387	387	288	288	288	288	288	288
Panel(4): $W_1 = W_{news}$ **and** $W_2 = W_{IBES}$												
MG	0.318	-0.007	-0.013	-0.007	-0.006	-0.002	0.160	-0.006	-0.004	-0.003	-0.002	0.002
	(0.018)	(0.004)	(0.002)	(0.003)	(0.003)	(0.002)	(0.015)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
%sig	79.6%	36.2%	23.8%	20.9%	22.2%	19.4%	79.7%	27.6%	17.1%	17.6%	15.9%	19.1%
N_p	387	387	387	387	387	387	340	340	340	340	340	340
Panel(5): $W_1 = W_{news}$ **and** $W_2 = W_{Customer-Supplier}$												
MG	0.418	-0.014	-0.017	-0.009	-0.008	-0.001	0.053	0.002	-0.002	-0.004	0.005	0.002
	(0.018)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.014)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)
%sig	84.2%	45.0%	37.2%	27.1%	27.4%	18.9%	71.2%	15.3%	13.6%	11.9%	6.8%	15.3%
N_p	387	387	387	387	387	387	59	59	59	59	59	59
Panel(6): $W_1 = W_{news}$ **and** $W_2 = W_{Geographic}$												
MG	0.399	-0.007	-0.014	-0.003	-0.004	-0.001	0.074	-0.009	-0.005	-0.008	-0.008	-0.000
	(0.018)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.016)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
%sig	83.7%	33.6%	24.8%	23.5%	21.4%	16.5%	59.3%	25.7%	19.2%	13.0%	12.7%	14.3%
N_p	387	387	387	387	387	387	307	307	307	307	307	307
Panel(7): $W_1 = W_{news}$ **and** $W_2 = W_{Union}$												
MG	0.292	0.002	-0.006	-0.000	0.001	-0.001	0.305	-0.016	-0.010	-0.010	-0.009	-0.001
	(0.016)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.018)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
%sig	76.0%	25.6%	18.1%	18.9%	19.4%	17.1%	83.0%	25.4%	17.5%	18.8%	17.0%	18.0%
N_p	387	387	387	387	387	387	394	394	394	394	394	394

Table 2.9 **QML estimation results of two-W spatial factor model with Fama-French five factors, the momentum factor, and the industry factor (Equation 2.37).**
Note: we only report spatial parameters here. W_{news} is constructed using threshold $m = 1$.

Table 2.8 and Table 2.9 shows the estimation results for Equation 2.36 and Equation 2.37, respectively. Although the magnitude of local dependencies among news-implied peers is weakened by the introduction of other networks, our new-implied links are still important channels of risk spillovers. The mean group (MG) estimates of $\psi_{1,0}$ are around $0.27 - 0.48$, with more than 70% of parameters being individually significant across different specifications.

Even if we condition on the union of all alternative linkages (i.e., $W_2 = W_{union}$), the magnitude of local dependencies among news-implied peers is still quite large. For the specification without industry factor, average $\psi_{2,0}$ is larger than average $\psi_{1,0}$. However, the introduction of industry factor reduces $\hat{\psi}_{2,0}^{MG}$ while leaving $\hat{\psi}_{1,0}^{MG}$ unchanged, making the local spillover effect via news-implied network equally strong as the effect via the union of all other alternative networks. The results confirm that the novel dataset carries additional information on top of existing networks. The statistically and economically significant local dependencies among the news-implied peers cannot be explained by other existing networks.

2.7 Conclusion

This paper studies a heterogeneous coefficient spatial factor model, which addresses both strong cross-sectional dependence and a very flexible form of weak cross-sectional dependence in equity returns. Theoretically, it extends classical asset pricing models like CAPM and APT, which only consider the strong form of cross-sectional dependence. We characterize how local dependence affects asset returns under the assumption of no asymptotic arbitrage. Empirically, we focus on the weak/local dependency in equity returns, which is an area less explored in empirical financial studies due to data availability issues. Utilizing the novel business news-implied linkage data, we construct the channels through which the local shocks transmit. We adopt a flexible heterogeneous coefficient spatial-temporal model, and we find that stocks linked via business news co-mentioning exhibit excess co-movement beyond that is predicted by standard asset pricing models like CAPM and APT. Exposures to common risk factors and local interactions are two distinct mechanism that jointly explain the co-movement in asset returns. It is important for investors and policy makers to separately analyse the two types of dependencies to fully understand what type of risk are they exposed to.

One interesting question for future work is whether the spatial factor model can be applied for portfolio construction problem. With the presence of both factor-driven strong dependence and the remaining weak dependence, literature on high-dimensional equity returns covariance matrix usually consider the following estimator

$$\hat{\Sigma}_y = \hat{\mathbf{B}} \hat{cov}(\boldsymbol{f})_t \hat{\mathbf{B}}' + \hat{\Sigma}_\varepsilon, \tag{2.38}$$

where $\hat{\Sigma}_\varepsilon$ is a regularised sparse error covariance matrix, and the estimation of Σ_y is achieved in two steps. The spatial factor model we study in this paper implies the following covariance

structure

(2.39)

which can be estimated in a single step.

Another interesting future work is the formal testing of no arbitrage $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ for the spatial factor model when N is large. Pesaran and Yamagata (2012), Pesaran and Yamagata (2017) consider testing for alpha in factor models with large N, how to extend the theory is non-trivial and needs thorough analysis.

Appendices

(A) Proofs of Theorems and Corollaries

(A.1)Proof of Theorem 1

This proof is heavily borrowed from Kou et al. (2018) and Ingersoll Jr (1984). Under Assumption 3, $(I-\Psi W)$ is invertible and we denote the inverse as $G(\psi)=(I-\Psi W)^{-1}$. We rewrite the spatial factor model (Equation 2.7) as

$$\boldsymbol{r}=G\boldsymbol{\alpha}+G\mathbf{B}\mathbf{f}+G\boldsymbol{\varepsilon}. \tag{40}$$

We let

$$\dot{\boldsymbol{\alpha}}=G\boldsymbol{\alpha},\ \ \dot{\mathbf{B}}=G\mathbf{B},\ \ \dot{\boldsymbol{\varepsilon}}=G\boldsymbol{\varepsilon}. \tag{41}$$

The spatial factor model can be written as a reduced-form factor model

$$\boldsymbol{r}=\dot{\boldsymbol{\alpha}}+\dot{\mathbf{B}}\mathbf{f}+\dot{\boldsymbol{\varepsilon}}. \tag{42}$$

In particular, the covariance matrix of the reduced form error is

$$\dot{\Omega}=E(\dot{\boldsymbol{\varepsilon}}\dot{\boldsymbol{\varepsilon}}')=G\Omega G'. \tag{43}$$

We follow Ingersoll Jr (1984), and factor the positive definite covariance matrix $\dot{\Omega}$ as $\dot{\Omega}=CC'$, where C is a nonsingular matrix. Now consider a subsequences of assets. For the nth economy, consider the orthogonal projection of the vector $(C^n)^{-1}\dot{\boldsymbol{\alpha}}^n$ into the space spanned by $(C^n)^{-1}\mathbf{1}^n$ and the columns of $(C^n)^{-1}\dot{\mathbf{B}}^n$ as follows:

$$(C^n)^{-1}\dot{\boldsymbol{\alpha}}^n=(C^n)^{-1}\mathbf{1}^n\lambda_0^n+(C^n)^{-1}\dot{\mathbf{B}}^n\boldsymbol{\lambda}^n+\boldsymbol{u}^n. \tag{44}$$

By the nature of orthogonal projection,

$$0=(\dot{\mathbf{B}}^n)'((C^n)')^{-1}\boldsymbol{u}^n=(\mathbf{1}^n)'((C^n)')^{-1}\boldsymbol{u}^n. \tag{45}$$

Given the pricing error $\boldsymbol{v}^n$ we defined in Equation 2.16, the reduced form pricing error from the reduced form factor model Equation 42 is $\dot{\boldsymbol{v}}^n=G\boldsymbol{v}^n$

$$\dot{\boldsymbol{v}}^n=\dot{\boldsymbol{\alpha}}^n-\mathbf{1}^n\lambda_0^n-\dot{\mathbf{B}}^n\boldsymbol{\lambda}^n=G^n(\boldsymbol{\alpha}^n-(G^n)^{-1}\mathbf{1}^n\lambda_0^n-\mathbf{B}^n\boldsymbol{\lambda}^n)=G^n\boldsymbol{v}^n. \tag{46}$$

The reduced form pricing error $\dot{\boldsymbol{v}}^n$ can be written as $\dot{\boldsymbol{v}}^n = C^n \boldsymbol{u}^n$ directly from Equation 55. Using the orthogonal conditions Equation 56 and the factorization $\dot{\Omega}^n = C^n (C^n)'$, we have:

$$(\dot{\mathbf{B}}^n)'(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{v}}^n = (\mathbf{1}^n)'(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{v}}^n = 0. \tag{47}$$

Consider a zero cost portfolio $\boldsymbol{c}^n = (\dot{\Omega}^n)^{-1}\dot{\boldsymbol{v}}^n[(\dot{\boldsymbol{v}}^n)'(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{v}}^n]^{-1}$

$$(\mathbf{1}^n)'\boldsymbol{c}^n = (\mathbf{1}^n)'(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{v}}^n[(\dot{\boldsymbol{v}}^n)'(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{v}}^n]^{-1} = 0, \tag{48}$$

with expected return

$$E((\boldsymbol{c}^n)'\boldsymbol{r}^n) = (\boldsymbol{c}^n)'\dot{\boldsymbol{\alpha}}^n = [(\dot{\boldsymbol{v}}^n)'(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{v}}^n]^{-1}(\dot{\boldsymbol{v}}^n)'(\dot{\Omega}^n)^{-1}(\mathbf{1}^n\lambda_0^n + \dot{\mathbf{B}}^n\boldsymbol{\lambda}^n + \dot{\boldsymbol{v}}^n) = 1, \tag{49}$$

and variance

$$\begin{aligned} Var((\boldsymbol{c}^n)'\boldsymbol{r}^n) &= (\boldsymbol{c}^n)'Var(\boldsymbol{r}^n)\boldsymbol{c}^n \\ &= [(\dot{\boldsymbol{v}}^n)'(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{v}}^n]^{-1}(\dot{\boldsymbol{v}}^n)'(\dot{\Omega}^n)^{-1}(\dot{\mathbf{B}}^n(\dot{\mathbf{B}}^n)' + \dot{\Omega}^n)(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{v}}^n[(\dot{\boldsymbol{v}}^n)'(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{v}}^n]^{-1} \\ &= [(\dot{\boldsymbol{v}}^n)'(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{v}}^n]^{-1} = [(G^n\boldsymbol{v}^n)'(G^n\Omega^n(G^n)')^{-1}G^n\boldsymbol{v}^n]^{-1} \\ &= [(\boldsymbol{v}^n)'(\Omega^n)^{-1}(\boldsymbol{v}^n)]^{-1} \end{aligned} \tag{50}$$

If the weighted sum of squared pricing errors $(\boldsymbol{v}^n)'(\Omega^n)^{-1}(\boldsymbol{v}^n)$ is not uniformly bounded (i.e., Equation 2.17 is violated), then the variance of this portfolio would go to zero along some subsequence, and the asymptotic arbitrage opportunity described in Equation 2.14 exists.

(A.2)Proof Corollary 1.1

This is a direct result of Theorem 3 from Chamberlain and Rothschild (1983).

(A.3)Proof Corollary 1.2

Again, we look at the reduced form factor model

$$\boldsymbol{r} = \dot{\boldsymbol{\alpha}} + \dot{\mathbf{B}}\mathbf{f} + \dot{\boldsymbol{\varepsilon}}, \tag{51}$$

where

$$\dot{\boldsymbol{\alpha}} = G\boldsymbol{\alpha}, \;\; \dot{\mathbf{B}} = G\mathbf{B}, \;\; \dot{\boldsymbol{\varepsilon}} = G\boldsymbol{\varepsilon}. \tag{52}$$

$$\dot{\Omega} = E(\dot{\boldsymbol{\varepsilon}}\dot{\boldsymbol{\varepsilon}}') = G\Omega G'. \tag{53}$$

The risk factors are given by

$$f_k = \tilde{f}_k - E(\tilde{f}_k) \text{ for } k = 1, \ldots, K. \tag{54}$$

We factor the positive definite covariance matrix $\dot{\Omega}$ as $\dot{\Omega} = CC'$, where C is a nonsingular matrix. Now consider a subsequences of assets. For the nth economy, consider the orthogonal projection of the vector $(C^n)^{-1}(\dot{\boldsymbol{\alpha}}^n - r_f \mathbf{1}^n)$ onto the space spanned by columns of $(C^n)^{-1}\dot{\mathbf{B}}^n$ as follows:

$$(C^n)^{-1}(\dot{\boldsymbol{\alpha}}^n - r_f \mathbf{1}^n) = (C^n)^{-1}\dot{\mathbf{B}}^n \boldsymbol{\lambda}^n + \boldsymbol{u}^n. \tag{55}$$

By the nature of orthogonal projection,

$$(\dot{\mathbf{B}}^n)'((C^n)')^{-1}\boldsymbol{u}^n = 0. \tag{56}$$

We define

$$\boldsymbol{v}^n = \boldsymbol{\alpha}^n - r_f(I_n - \Psi^n W^n)\mathbf{1}^n - \mathbf{B}^n E(\tilde{\boldsymbol{f}}) = \boldsymbol{\alpha}^n - r_f(G^n)^{-1}\mathbf{1}^n - \mathbf{B}^n E(\tilde{\boldsymbol{f}}). \tag{57}$$

Define the the reduced form pricing error

$$\dot{\boldsymbol{v}}^n = \dot{\boldsymbol{\alpha}}^n - r_f \mathbf{1}^n - \dot{\mathbf{B}}^n \boldsymbol{\lambda}^n. \tag{58}$$

Under the assumption that factors are traded, $\boldsymbol{\lambda}^n = E(\tilde{\boldsymbol{f}})$. And we have $\dot{\boldsymbol{v}}^n = G^n \boldsymbol{v}^n$ as:

$$\dot{\boldsymbol{v}}^n = \dot{\boldsymbol{\alpha}}^n - r_f \mathbf{1}^n - \dot{\mathbf{B}}^n \boldsymbol{\lambda}^n = G^n(\boldsymbol{\alpha}^n - r_f(G^n)^{-1}\mathbf{1}^n - \mathbf{B}^n E(\tilde{\boldsymbol{f}})) = G^n \boldsymbol{v}^n. \tag{59}$$

Given there exists risk free asset, consider a zero-cost portfolio which take a long position $\boldsymbol{c}^n$ in the risky assets and short position $(\boldsymbol{c}^n)'\mathbf{1}^n$ in the risk free asset, where $\boldsymbol{c}^n = (\dot{\Omega}^n)^{-1}\dot{\boldsymbol{v}}^n[(\dot{\boldsymbol{v}}^n)'(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{v}}^n]^{-1}$. This portfolio generates expected return

$$\begin{aligned}
&E((\boldsymbol{c}^n)'(\boldsymbol{r}^n - r_f \mathbf{1}^n)) \\
&= E((\boldsymbol{c}^n)'(\dot{\boldsymbol{\alpha}}^n - r_f \mathbf{1}^n)) + E((\boldsymbol{c}^n)'\dot{\mathbf{B}}^n \boldsymbol{f}) + E((\boldsymbol{c}^n)'\dot{\boldsymbol{\varepsilon}}^n) \\
&= E((\boldsymbol{c}^n)'(\dot{\boldsymbol{\alpha}}^n - r_f \mathbf{1}^n)) = E((\boldsymbol{c}^n)'(\dot{\mathbf{B}}^n \boldsymbol{\lambda}^n + \dot{\boldsymbol{v}}^n)) \\
&= [(\dot{\boldsymbol{v}}^n)'(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{v}}^n]^{-1}(\dot{\boldsymbol{v}}^n)'(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{v}}^n = 1,
\end{aligned} \tag{60}$$

and variance

$$\begin{aligned}
&Var((\boldsymbol{c}^n)'(\boldsymbol{r}^n - r_f\mathbf{1}^n)) \\
&= (\boldsymbol{c}^n)'Var(\dot{\mathbf{B}}^n\boldsymbol{f} + \dot{\boldsymbol{\varepsilon}}^n)\boldsymbol{c}^n \\
&= [(\dot{\boldsymbol{\nu}}^n)'(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{\nu}}^n]^{-1}(\dot{\boldsymbol{\nu}}^n)'(\dot{\Omega}^n)^{-1}(\dot{\mathbf{B}}^n(\dot{\mathbf{B}}^n)' + \dot{\Omega}^n)(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{\nu}}^n[(\dot{\boldsymbol{\nu}}^n)'(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{\nu}}^n]^{-1} \\
&= [(\dot{\boldsymbol{\nu}}^n)'(\dot{\Omega}^n)^{-1}\dot{\boldsymbol{\nu}}^n]^{-1} = [(G^n\boldsymbol{\nu}^n)'(G^n\Omega^n(G^n)')^{-1}G^n\boldsymbol{\nu}^n]^{-1} \\
&= [(\boldsymbol{\nu}^n)'(\Omega^n)^{-1}(\boldsymbol{\nu}^n)]^{-1}
\end{aligned} \tag{61}$$

For pricing errors $\boldsymbol{\nu}^n = \boldsymbol{\alpha}^n - r_f(I_n - \Psi^n W^n)\mathbf{1}^n - \mathbf{B}^n E(\tilde{\boldsymbol{f}})$ defined in Equation 57, if the weighted sum of squared pricing errors $(\boldsymbol{\nu}^n)'(\Omega^n)^{-1}(\boldsymbol{\nu}^n)$ is not uniformly bounded (i.e., Equation 2.17 is violated), then the variance of this portfolio would go to zero along some subsequence, and the asymptotic arbitrage opportunity described in Equation 2.14 exists.

When there exists risk free rate r_f, and if we write the risk factor f_k before de-meaning as $\tilde{f}_k$, then the spatial factor model (Equation 2.7) can be written as:

$$\begin{aligned}
&\tilde{\boldsymbol{r}} = \tilde{\boldsymbol{\alpha}} + \mathbf{B}\tilde{\mathbf{f}} + \Psi W\tilde{\boldsymbol{r}} + \boldsymbol{\varepsilon}, \\
&\text{where } \tilde{\boldsymbol{r}} = \boldsymbol{r} - r_f\mathbf{1} \text{ is the vector of excess returns,} \\
&\tilde{\mathbf{f}} = (\tilde{f}_1, \ldots, \tilde{f}_K)', \\
&\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} - (I - \Psi W)\mathbf{1}r_f - \mathbf{B}E(\tilde{\mathbf{f}}).
\end{aligned} \tag{62}$$

Comparing this pricing errors $\boldsymbol{\nu}^n = \boldsymbol{\alpha}^n - r_f(I_n - \Psi^n W^n)\mathbf{1}^n - \mathbf{B}^n E(\tilde{\boldsymbol{f}})$ with Equation 62, we can tell $\boldsymbol{\nu}^n = \tilde{\boldsymbol{\alpha}}^n$, and the asymptotic no arbitrage condition is equivalent to

$$\tilde{\boldsymbol{\alpha}}^n \approx 0. \tag{63}$$

Comparing $\boldsymbol{\nu}^n = \tilde{\boldsymbol{\alpha}}^n = \boldsymbol{\alpha}^n - r_f(I_n - \Psi^n W^n)\mathbf{1}^n - \mathbf{B}^n E(\tilde{\boldsymbol{f}})$ with Equation 2.16, we can tell

$$\begin{aligned}
&\lambda_0^n = r_f, \\
&\boldsymbol{\lambda}^n = E(\tilde{\mathbf{f}}).
\end{aligned} \tag{64}$$

(A.4)Proof Corollary 1.3

We first rewrite the spatial-factor model with the dependent variable being the excesss returns (Equation 62) as

$$(I - \Psi W)\tilde{\boldsymbol{r}} = G^{-1}\tilde{\boldsymbol{r}} = \tilde{\boldsymbol{\alpha}} + \mathbf{B}\tilde{\mathbf{f}} + \boldsymbol{\varepsilon}. \tag{65}$$

Suppose the asset returns from an infinite economy are generated by:

$$(G^n)^{-1}\tilde{\boldsymbol{r}}^n = \tilde{\boldsymbol{\alpha}}^n + \mathbf{B}^n\tilde{\mathbf{f}} + \boldsymbol{\varepsilon}^n \tag{66}$$

For any fixed $\delta > 0$, assume $I(\mid \tilde{\alpha}_j^n \mid > \delta) = 1$ for $j = 1,\ldots,N(n,\delta)$. For each of those $N(n,\delta)$ elements, we can construct a zero cost portfolio as following way.

Take the jth element for example. Denote the jth column of Identify matrix I_n by $\boldsymbol{e}_j$. If $\tilde{\alpha}_j^n > \delta$, consider a zero-cost portfolio which takes a long position $\boldsymbol{e}_j'(G^n)^{-1}$ in excess returns $\tilde{\boldsymbol{r}}^n$, and short position $\boldsymbol{e}_j'\mathbf{B}^n$ in the zero-cost traded factors $\tilde{\mathbf{f}}$. If $\tilde{\alpha}_j^n < -\delta$, consider a zero-cost portfolio which takes a short position $\boldsymbol{e}_j'(G^n)^{-1}$ in excess returns $\tilde{\boldsymbol{r}}^n$, and long position $\boldsymbol{e}_j'\mathbf{B}^n$ in the zero-cost traded factors $\tilde{\mathbf{f}}$. The portfolio is a zero-cost one because the long and short position are both zero-cost. The portfolio has expected return $\mid \tilde{\alpha}_j^n \mid > \delta$, and variance $\sigma_j^2 < \bar{\sigma}^2$.

We can construct $N(n,\delta)$ such portfolios. Consider a new portfolio that takes equal weight in these $N(n,\delta)$ portfolios. This new portfolio is zero-cost, with expected return $\frac{\sum_{j=1}^{N(n,\delta)}|\tilde{\alpha}_j^n|}{N(n,\delta)} > \delta > 0$. For uncorrelated errors, the variance of this portfolio is smaller than $\frac{\bar{\sigma}^2}{N(n,\delta)}$. If Equation 2.23 fails, and $N(n,\delta)$ is diverging, then we have asymptotic arbitrage.

(B)Identification and Inference of the Heterogeneous Spatial-Temporal Model

Aquaro et al. (2020) studies the conditions under which $\boldsymbol{\theta}_0$ is identified, and establishes consistency and asymptotic normality of the estimator. Write the $(N*(K+L+3))$ by 1 vector $\boldsymbol{\theta} = (\tilde{\boldsymbol{\alpha}}', \boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_K', \boldsymbol{\psi}_0', \ldots, \boldsymbol{\psi}_L', \boldsymbol{\sigma}_{\varepsilon^2}')' = (\boldsymbol{b}', \boldsymbol{\psi}_0', \boldsymbol{\sigma}_{\varepsilon^2}')'$, where $\boldsymbol{b} = (\tilde{\boldsymbol{\alpha}}', \boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_K', \boldsymbol{\psi}_1' \ldots, \boldsymbol{\psi}_L')'$ is a $(N*(K+L+1))$ by 1 vector that contains all the parameters associated with weakly exogenous variables $\boldsymbol{x}_t$. The following assumptions are made:

Assumption 5. *The parameter vector $\boldsymbol{\theta} = (\boldsymbol{b}', \boldsymbol{\psi}_0', \boldsymbol{\sigma}_{\varepsilon^2}')'$ belongs to $\Theta = \Theta_b \times \Theta_{\boldsymbol{\psi}_0} \times \Theta_\sigma \subset R^{N*(K+L+1)} \times R^N \times R^N$, a subset of the $(N*(K+L+3))$ dimensional Euclidean space $\mathbb{R}^{N*(K+L+3)}$. Θ is a closed and bounded (compact) set, and $\boldsymbol{\theta}_0$ is an interior point of Θ.*

Assumption 6. *The error terms $\{\varepsilon_{it}, i = 1,\ldots,N; t = 1,\ldots,T\}$ are independently distributed over i and t. For filtration $\mathscr{F}_t = (\boldsymbol{x}_t, \boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \ldots)$, $E(\varepsilon_{it} \mid \mathscr{F}_t) = 0$, $E(\varepsilon_{it}^2 \mid \mathscr{F}_t) = \sigma_{i0}^2$, for $i =$*

$1,\ldots,N$, *so there is no conditional heteroskedasticity.* $inf_i\sigma_{i0}^2 > c > 0$ *and* $sup_i\sigma_{i0}^2 < \bar{\sigma}^2 < \infty$, *and* $E(\mid \varepsilon_{it} \mid^p \mid \mathscr{F}_t) = E((\mid \varepsilon_{it} \mid^p) = \bar{\omega}_{ip} < \bar{c}$, *for all* i *and* t, *where* $1 \leqslant p \leqslant 4+\varepsilon$, *for some* $\varepsilon > 0$.

Assumption 7. *(a)* $\boldsymbol{x}_t$ *are stationary processes, that satisfy the moment condition* $sup_{i,t,l}E(\mid x_{it,l} \mid^{2+g}) < \bar{c}$, *for some* $g > 0$, $i = 1,\ldots,N$, $t = 1,\ldots,T$, $l = 1,\ldots,(K+L+1)$.
(b) $E(\boldsymbol{x}_t\boldsymbol{x}_t') = \Sigma_{xx}$, *where entry* $\Sigma_{ij} = E(\boldsymbol{x}_{it}\boldsymbol{x}_{jt}')$ *exists for all* i *and* j, *such as* $sup_{i,j}\|\Sigma_{ij}\| < \bar{c}$, *and* Σ_{ii} *is a* $k \times k$ *non-singular matrix with* $inf_i[\lambda_{min}(\Sigma_{ii})] > c > 0$, *and* $sup_i[\lambda_{min}(\Sigma_{ii})] < \bar{c} < \infty$.
(c) $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t' \xrightarrow{a.s} \Sigma_{xx}$ *as* $T \to \infty$.

Assumption 8. *(a) The adjacency matrix* W *is known, with zeros on the diagonal.*
(b) The adjacency matrix W *has bounded row sum norm, and* $\|W\|_\infty < c < \infty$, *and*

$$\sup_{\psi_i \in \Theta_{\boldsymbol{\psi}}} \mid \psi_i \mid < \frac{1}{\|W\|_\infty}. \tag{67}$$

Definition 1. The set $\mathscr{N}_c(\boldsymbol{\sigma}_0^2)$ in the closed neighbourhood of $\boldsymbol{\sigma}_0^2$ if:

$$\mathscr{N}_c(\boldsymbol{\sigma}_0^2) = \{\boldsymbol{\sigma}_0^2 \in \Theta_\sigma, \mid \sigma_{i0}^2/\sigma_i^2 - 1 \mid < c_i, \text{for } i = 1,\ldots,N\}, \tag{68}$$

for some $c_i > 0$, where Θ_σ is a compact subset of $\mathscr{R}^N$.

Assumption 9. *The* $(N*(K+L+3))$ *by* 1 *vector* $\boldsymbol{\theta} = (\boldsymbol{b}', \boldsymbol{\psi}_0', \boldsymbol{\sigma}_{\varepsilon^2}')'$ *belongs to* $\Theta_c = \Theta_b \times \Theta_{\boldsymbol{\psi}_0} \times \mathscr{N}_c(\boldsymbol{\sigma}_0^2)$. Θ_b *and* $\Theta_{\boldsymbol{\psi}_0}$ *are compact subsets of* $\mathscr{R}^{N*(K+L+1)}$ *and* $\mathscr{R}^N$, *respectively, and* $\mathscr{N}_c(\boldsymbol{\sigma}_0^2)$ *is defined in Definition 1, and* Θ_c *is a subset of the* $(N*(K+L+3))$ *dimensional Euclidean space,* $\mathscr{R}^{N*(K+L+3)}$.

The identification results are given by the following proposition:

Proposition 2. *Suppose that Assumptions 1-5 hold, consider a heterogeneous coefficient spatial-temporal model given by Equation 2.25 and log-likelihood function given by Equation 2.26. For fixed* N, K *and* L, *the* $(N*(K+L+3))$ *dimensional true parameter vector* $\boldsymbol{\theta}_0$ *is almost surely locally identified on* Θ_c.

The main inference results are given by the following proposition:

Proposition 3. *Suppose that Assumptions 1-5 hold, consider a heterogeneous coefficient spatial-temporal model given by Equation 2.25. For fixed N, K and L, the* $(N*(K+L+3))$ *dimensional QML estimator of* $\boldsymbol{\theta}_0$ *is denoted as* $\hat{\boldsymbol{\theta}}_{QMLE}$*, which is almost surely locally consistent for* $\boldsymbol{\theta}_0$ *on* Θ_c*, and has the following asymptotic distribution:*

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_{QMLE} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{V}_\theta), \tag{69}$$

where $\boldsymbol{V}_\theta$ *is the asymptotic covariance matrix, which has a standard sandwich form:*

$$\boldsymbol{V}_\theta = \boldsymbol{H}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{J}(\boldsymbol{\theta}_0, \gamma)\boldsymbol{H}^{-1}(\boldsymbol{\theta}_0), \tag{70}$$

where $\boldsymbol{H}(\boldsymbol{\theta}_0) = \lim_{T\to\infty} E_0(-\frac{1}{T}\ell_T(\boldsymbol{\theta})\boldsymbol{\theta}\boldsymbol{\theta}')$ *is the Hessian, and* $\boldsymbol{J}(\boldsymbol{\theta}_0, \gamma)$ *is the asymptotic variance of the score, which depends on the distribution of the errors. In the case of Gaussian errors,* $\gamma = 2$*, and* $\boldsymbol{H}(\boldsymbol{\theta}_0) = \boldsymbol{J}(\boldsymbol{\theta}_0, 2)$.

Remark: Proposition 1 and Proposition 2 describe the identification results and asymptotic distribution for each individual parameter in the $(N*(K+L+3))$ by 1 vector. When $T \to \infty$, estimation and inference can be conducted for any N.

(C)Supplementary Figures and Tables

Number of unique news stories	88,316,898
Number of stories remaining after removing topics including analyst recommendations, ratings changes, and index movements	87,841,641
Of these:	
Number of stories tag sample companies	8,341,848
Of these:	
Number of stories that mention only one company	5,507,772 (66.03%)
Number of stories that mention exactly two companies	1,637,256 (19.63%)
Number of stories that mention more than two companies	1,196,820 (14.34%)

Table 10 Descriptive statistics for RavenPack Equity files Dow Jones Edition for the period January 2004 to December 2015.

Number of yearly window a pair gets identified	Frequency	Percentage	Cumulative Percentage
0	217024	72.80%	72.80%
1	40178	13.48%	86.28%
2	13302	4.46%	90.74%
3	7116	2.39%	93.13%
4	4522	1.52%	94.65%
5	3236	1.09%	95.74%
6	2506	. 0.84%	96.58%
7	2022	0.68%	97.26%
8	1804	0.61%	97.87%
9	1508	0.51%	98.38%
10	1350	0.45%	98.83%
11	1232	0.41%	99.24%
12	2316	0.78%	100%

Table 11 Frequency distribution table of the number of yearly link identification windows that a pair gets identified as economic neighbours for all possible pairs (i, j) in our sample.
Note: A pair identified in k yearly windows could get multiple co-mentions within each window.

	Finance	Durbs	Energy	Hi-tec	Health	Manuf	Nondur	Other	Shops	Tel	Utilities
Finance	1840	81	256	777	315	529	273	573	568	116	235
	0.33	0.01	0.05	0.14	0.06	0.10	0.05	0.10	0.10	0.02	0.04
Durbs	81	12	14	67	16	72	27	49	45	10	13
	0.20	0.03	0.03	0.17	0.04	0.18	0.07	0.12	0.11	0.02	0.03
Energy	256	14	372	147	42	115	51	153	83	20	172
	0.18	0.01	0.26	0.10	0.03	0.08	0.04	0.11	0.06	0.01	0.12
Hi-tec	777	67	147	1376	227	419	182	439	403	126	86
	0.18	0.02	0.03	0.32	0.05	0.10	0.04	0.10	0.09	0.03	0.02
Health	315	16	42	227	370	111	71	134	143	28	19
	0.21	0.01	0.03	0.15	0.25	0.08	0.05	0.09	0.10	0.02	0.01
Manuf	529	72	115	419	111	470	134	287	211	43	62
	0.220	0.03	0.05	0.17	0.05	0.19	0.05	0.12	0.09	0.02	0.03
Nondur	273	27	51	182	71	134	196	152	244	42	25
	0.20	0.02	0.04	0.13	0.05	0.10	0.14	0.11	0.17	0.03	0.02
Other	573	49	153	439	134	287	152	344	295	63	138
	0.22	0.02	0.06	0.17	0.05	0.11	0.06	0.13	0.11	0.02	0.05
Shops	568	45	83	403	143	211	244	295	698	73	40
	0.20	0.02	0.03	0.14	0.05	0.08	0.09	0.11	0.25	0.03	0.01
Telcm	116	10	20	126	28	43	42	63	73	18	22
	0.21	0.02	0.04	0.22	0.05	0.08	0.07	0.11	0.13	0.03	0.04
Utilities	235	13	172	86	19	62	25	138	40	22	366
	0.20	0.01	0.15	0.07	0.02	0.05	0.02	0.12	0.03	0.02	0.31

Table 12 **Links aggregated at industry level**. Note: The adjacency matrix is construct using threshold $m = 1$. we use Fama-French 12 industry classification. For each panel, the first row gives the number of intra or inter industry pairs indentified, and the second gives the proportion to total number of links firms in that industry have.

	Finance	Durbs	Energy	Hi-tec	Health	Manuf	Nondur	Other	Shops	Tel	Utilities
Finance	1496	65	193	566	233	377	193	451	397	84	173
	0.35	0.02	0.05	0.13	0.06	0.09	0.05	0.11	0.09	0.02	0.04
Durbs	65	12	8	41	8	43	14	36	29	8	7
	0.24	0.04	0.03	0.15	0.03	0.16	0.05	0.13	0.11	0.03	0.03
Energy	193	8	294	87	18	61	29	95	44	11	110
	0.20	0.01	0.31	0.09	0.02	0.06	0.03	0.10	0.05	0.01	0.12
Hi-tec	566	41	87	1040	123	254	103	311	254	85	40
	0.19	0.01	0.03	0.36	0.04	0.09	0.04	0.11	0.09	0.03	0.01
Health	233	8	18	123	288	64	40	86	75	17	4
	0.24	0.01	0.02	0.13	0.30	0.07	0.04	0.09	0.08	0.02	0
Manuf	377	43	61	254	64	264	84	205	109	26	20
	0.25	0.03	0.04	0.17	0.04	0.18	0.06	0.14	0.07	0.02	0.01
Nondur	193	14	29	103	40	84	144	97	158	30	12
	0.21	0.02	0.03	0.11	0.04	0.09	0.16	0.11	0.17	0.03	0.01
Other	451	36	95	311	86	205	97	256	177	52	84
	0.24	0.02	0.05	0.17	0.05	0.11	0.05	0.14	0.10	0.03	0.05
Shops	397	29	44	254	75	109	158	177	536	48	19
	0.22	0.02	0.02	0.14	0.04	0.06	0.09	0.10	0.29	0.03	0.01
Telcm	84	8	11	85	17	26	30	52	48	18	13
	0.21	0.02	0.03	0.22	0.04	0.07	0.08	0.13	0.12	0.05	0.03
Utilities	173	7	110	40	4	20	12	84	19	13	290
	0.22	0.01	0.14	0.05	0.01	0.03	0.02	0.11	0.02	0.02	0.38

Table 13 **Links aggregated at industry level**. Note: The adjacency matrix is construct using threshold $m = 2$. we use Fama-French 12 industry classification. For each panel, the first row gives the number of intra or inter industry pairs indentified, and the second gives the proportion to total number of links firms in that industry have.

	Finance	Durbs	Energy	Hi-tec	Health	Manuf	Nondur	Other	Shops	Tel	Utilities
Finance	1250	50	153	415	187	289	160	380	315	61	136
	0.37	0.01	0.05	0.12	0.06	0.09	0.05	0.11	0.09	0.02	0.04
Durbs	50	8	7	29	5	31	10	30	20	5	4
	0.25	0.04	0.04	0.15	0.03	0.16	0.05	0.15	0.10	0.03	0.02
Energy	153	7	246	54	11	42	19	72	22	8	60
	0.22	0.01	0.35	0.08	0.02	0.06	0.03	0.10	0.03	0.01	0.09
Hi-tec	415	29	54	832	82	172	63	235	164	73	23
	0.19	0.01	0.03	0.39	0.04	0.08	0.03	0.11	0.08	0.03	0.01
Health	187	5	11	82	246	44	26	67	44	9	2
	0.26	0.01	0.02	0.11	0.34	0.06	0.04	0.09	0.06	0.01	0
Manuf	289	31	42	172	44	186	55	156	62	16	10
	0.27	0.03	0.04	0.16	0.04	0.17	0.05	0.15	0.06	0.02	0.01
Nondur	160	10	19	63	26	55	112	75	114	21	5
	0.24	0.02	0.03	0.10	0.04	0.08	0.17	0.11	0.17	0.03	0.01
Other	380	30	72	235	67	156	75	210	126	30	63
	0.26	0.02	0.05	0.16	0.05	0.11	0.05	0.15	0.09	0.02	0.04
Shops	315	20	22	164	44	62	114	126	394	32	10
	0.24	0.02	0.02	0.13	0.03	0.05	0.09	0.10	0.30	0.02	0.01
Telcm	61	5	8	73	9	16	21	30	32	16	8
	0.22	0.02	0.03	0.26	0.03	0.06	0.08	0.11	0.11	0.06	0.03
Utilities	136	4	60	23	2	10	5	63	10	8	214
	0.25	0.01	0.11	0.04	0	0.02	0.01	0.12	0.02	0.01	0.40

Table 14 **Links aggregated at industry level**. Note: The adjacency matrix is construct using threshold $m = 3$. we use Fama-French 12 industry classification. For each panel, the first row gives the number of intra or inter industry pairs indentified, and the second gives the proportion to total number of links firms in that industry have.

	(1) factor component							(2) spatial-temporal component					
	α	β_1	β_2	β_3	β_4	β_5	β_6	ψ_0	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5
(1) Spatial CAPM													
MG	0.015	0.598						0.416	0.002	-0.008	0.001	-0.003	0.004
	(0.001)	(0.021)						(0.019)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	8.4%	91.6%						88.9%	51.2%	27.6%	20.9%	29.5%	21.2%
N_p	394	394						387	387	387	387	387	387
(2) Spatial APT (Fama-French three factors)													
MG	0.014	0.570	0.129	-0.127				0.450	0.002	-0.008	-0.001	-0.001	0.002
	(0.001)	(0.021)	(0.014)	(0.022)				(0.019)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	7.1%	90.1%	75.1%	81.5%				89.1%	52.7%	27.4%	20.4%	28.9%	19.1%
N_p	394	394	394	394				387	387	387	387	387	387
(3) Spatial APT (Fama-French five factors)													
MG	0.011	0.586	0.144	-0.127	0.142	0.177		0.453	0.007	-0.007	-0.001	-0.002	0.002
	(0.001)	(0.021)	(0.014)	(0.023)	(0.022)	(0.021)		(0.018)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	5.8%	90.6%	74.4%	82.2%	75.1%	73.9%		88.9%	51.4%	27.4%	20.7%	27.4%	18.3%
N_p	394	394	394	394	394	394		387	387	387	387	387	387
(4) Spatial APT (Fama-French five factors plus Momentum)													
MG	0.012	0.593	0.146	-0.137	0.139	0.184	-0.022	0.444	0.006	-0.007	-0.001	-0.002	0.001
	(0.001)	(0.021)	(0.014)	(0.021)	(0.022)	(0.021)	(0.007)	(0.019)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	4.8%	90.1%	74.1%	79.9%	75.4%	74.4%	59.1%	87.6%	51.7%	27.6%	22.2%	27.9%	18.3%
N_p	394	394	394	394	394	394	394	387	387	387	387	387	387

Table 15 **QML estimation results of heterogeneous spatial-temporal model using full sample, with W constructed using threshold $m = 2$**

	(1) factor component							(2) spatial-temporal component					
	α	β_1	β_2	β_3	β_4	β_5	β_6	ψ_0	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5
(1) Spatial CAPM													
MG	0.015	0.628						0.392	0.002	-0.008	0.001	-0.003	0.004
	(0.001)	(0.021)						(0.019)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	8.9%	92.9%						87.8%	49.7%	27.9%	20.3%	29.2%	22.7%
N_p	394	394						384	384	384	384	384	384
(2) Spatial APT (Fama-French three factors)													
MG	0.014	0.606	0.128	-0.117				0.420	0.002	-0.008	-0.000	-0.000	0.002
	(0.001)	(0.021)	(0.014)	(0.022)				(0.018)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	7.6%	92.1%	75.1%	82.0%				87.8%	51.8%	27.3%	20.6%	29.7%	20.6%
N_p	394	394	394	394				384	384	384	384	384	384
(3) Spatial APT (Fama-French five factors)													
MG	0.012	0.621	0.144	-0.116	0.141	0.173		0.423	0.006	-0.006	-0.001	-0.002	0.001
	(0.001)	(0.020)	(0.014)	(0.023)	(0.022)	(0.021)		(0.018)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	6.3%	92.6%	73.3%	81.7%	75.1%	72.3%		87.5%	50.8%	28.4%	20.6%	27.9%	19.8%
N_p	394	394	394	394	394	394		384	384	384	384	384	384
(4) Spatial APT (Fama-French five factors plus Momentum)													
MG	0.012	0.629	0.145	-0.130	0.138	0.182	-0.027	0.414	0.005	-0.007	-0.001	-0.002	0.001
	(0.001)	(0.021)	(0.014)	(0.021)	(0.022)	(0.021)	(0.007)	(0.018)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	5.8%	92.1%	73.6%	78.7%	75.1%	73.3%	61.4%	87.0%	51.0%	28.4%	22.1%	28.4%	20.1%
N_p	394	394	394	394	394	394	394	384	384	384	384	384	384

Table 16 **QML estimation results of heterogeneous spatial-temporal model using full sample, with W constructed using threshold $m = 3$**

	(1) factor component								(2) spatial-temporal component						
	α	β_1	β_2	β_3	β_4	β_5	β_6	β_I	ψ_0	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5
(1) Spatial CAPM+ Industry factor														
MG	0.009	0.257						0.410	0.361	-0.014	-0.023	-0.008	-0.007	0.002
	(0.002)	(0.024)						(0.021)	(0.018)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
%sig	7.9%	82.5%						84.5%	86.6%	49.4%	44.4%	27.9%	29.5%	22.0%
N_p	394	394						394	387	387	387	387	387	387
(2) Spatial APT (Fama-French three factors)+ Industry factor														
MG	0.007	0.230	-0.122	-0.147				0.456	0.381	-0.018	-0.018	-0.008	-0.006	-0.000
	(0.002)	(0.025)	(0.018)	(0.018)				(0.026)	(0.018)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
%sig	7.9%	80.5%	72.8%	80.2%				81.7%	84.8%	51.2%	39.0%	28.4%	28.7%	20.2%
N_p	394	394	394	394				394	387	387	387	387	387	387
(3) Spatial APT (Fama-French five factors)+ Industry factor														
MG	0.006	0.246	-0.110	-0.155	0.107	0.198		0.449	0.386	-0.013	-0.017	-0.009	-0.008	-0.001
	(0.002)	(0.025)	(0.018)	(0.019)	(0.019)	(0.017)		(0.025)	(0.017)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
%sig	6.9%	80.2%	72.1%	79.7%	69.3%	71.1%		82.5%	85.5%	46.5%	37.7%	27.6%	27.9%	19.9%
N_p	394	394	394	394	394	394		394	387	387	387	387	387	387
(4) Spatial APT (Fama-French five factors plus Momentum)														
MG	0.006	0.257	-0.106	-0.148	0.103	0.194	-0.001	0.439	0.383	-0.013	-0.017	-0.009	-0.008	-0.001
	(0.002)	(0.026)	(0.018)	(0.018)	(0.019)	(0.017)	(0.007)	(0.025)	(0.018)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
%sig	6.6%	80.5%	72.3%	78.9%	68.8%	68.8%	54.6%	80.7%	85.5%	45.7%	38.2%	26.9%	27.4%	18.9%
N_p	394	394	394	394	394	394	394	394	387	387	387	387	387	387

Table 17 **QML estimation results of Industry factors augmented models Equation 2.33 using full sample, with W constructed using threshold $m = 2$**

	(1) factor component								(2) spatial-temporal component					
	α	β_1	β_2	β_3	β_4	β_5	β_6	β_I	ψ_0	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5
(1) Spatial CAPM+ Industry factor														
MG	0.009	0.282						0.414	0.340	-0.014	-0.023	-0.007	-0.007	0.002
	(0.002)	(0.024)						(0.021)	(0.017)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
%sig	7.9%	83.2%						84.3%	84.6%	49.7%	44.5%	27.6%	30.2%	22.7%
N_P	394	394						394	384	384	384	384	384	384
(2) Spatial APT (Fama-French three factors)+ Industry factor														
MG	0.008	0.259	-0.125	-0.138				0.459	0.354	-0.017	-0.017	-0.008	-0.006	-0.000
	(0.002)	(0.025)	(0.018)	(0.018)				(0.026)	(0.017)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
%sig	8.1%	82.5%	73.3%	80.7%				82.0%	83.3%	50.3%	39.3%	27.9%	28.9%	20.3%
N_p	394	394	394	394				394	384	384	384	384	384	384
(3) Spatial APT (Fama-French five factors)+ Industry factor														
MG	0.006	0.275	-0.113	-0.146	0.106	0.195		0.454	0.358	-0.013	-0.017	-0.009	-0.007	-0.001
	(0.002)	(0.025)	(0.018)	(0.019)	(0.019)	(0.017)		(0.025)	(0.017)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
%sig	7.1%	82.7%	72.3%	79.7%	69.8%	70.3%		82.0%	83.9%	47.1%	38.0%	28.1%	27.6%	20.6%
N_p	394	394	394	394	394	394		394	384	384	384	384	384	384
(4) Spatial APT (Fama-French five factors plus Momentum)														
MG	0.006	0.257	-0.106	-0.148	0.103	0.194	-0.001	0.439	0.383	-0.013	-0.017	-0.009	-0.008	-0.001
	(0.002)	(0.026)	(0.018)	(0.018)	(0.019)	(0.017)	(0.007)	(0.025)	(0.018)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
%sig	6.6%	80.5%	72.3%	78.9%	68.8%	68.8%	54.6%	80.7%	85.5%	45.7%	38.2%	26.9%	27.4%	18.9%
N_p	394	394	394	394	394	394	394	394	387	387	387	387	387	387

Table 18 **QML estimation results of Industry factors augmented models Equation 2.33 using full sample, with W constructed using threshold $m = 3$**

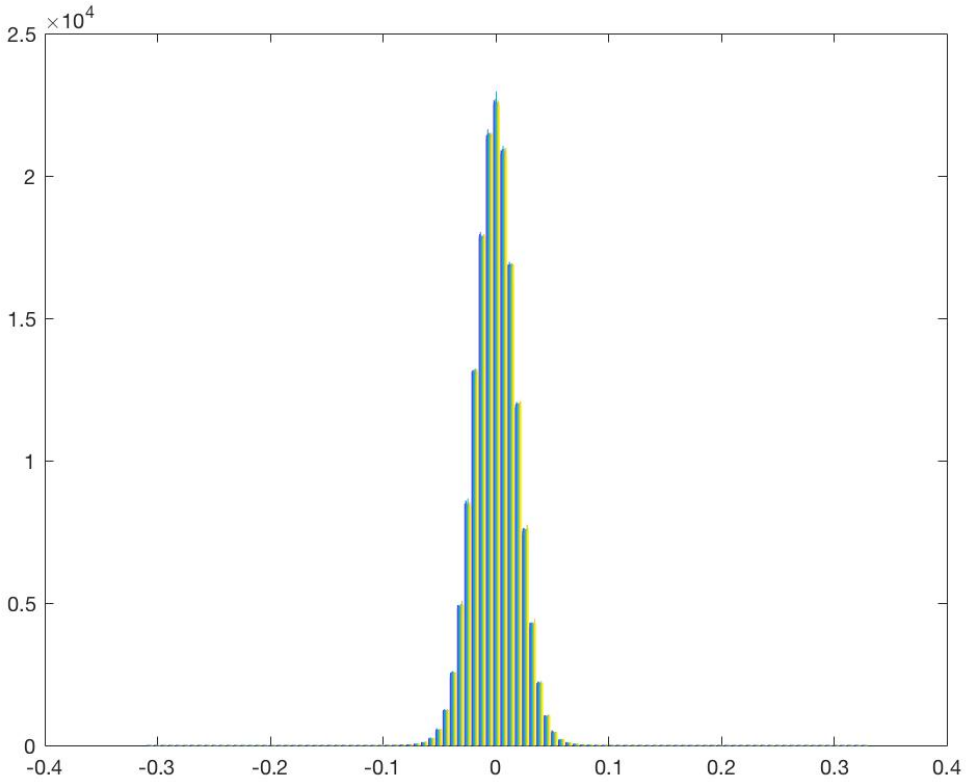


Figure 1 Histogram of bootstrapped $\hat{\rho}_{ij}^{b}$ for all $i \neq j, b = 1, \ldots, 500$.

Chapter 3

Dynamic Peer Groups of Arbitrage Characteristics

Abstract

We propose an asset pricing factor model constructed with semiparametric characteristics-based mispricing and factor loading functions. We approximate the unknown functions by B-splines sieve where the number of B-splines coefficients is diverging. We estimate this model and test the existence of the mispricing function by a power enhanced hypothesis test. The enhanced test solves the low power problem caused by diverging B-splines coefficients, with the strengthened power approaches to one asymptotically. We also investigate the structure of mispricing components through Hierarchical K-means Clusterings. We apply our methodology to CRSP (Center for Research in Security Prices) data for the US stock market with one-year rolling windows during 1967-2017. This empirical study shows the presence of mispricing functions in certain time blocks. We also find that distinct clusters of the same characteristics lead to similar arbitrage returns, forming a "peer group" of arbitrage characteristics.

Keywords: Semiparametric; Characteristics-based; Peer Groups; Power-enhanced test
JEL Classification: C14; G11; G12

3.1 Introduction

Stock returns have both common and firm-specific components. Ross (1976) proposed Arbitrage Pricing Theory (APT) to summarize that expected returns on financial assets can be modeled as a linear combination of risk factors. In such a model, each asset has a sensitivity beta to the risk factor. The APT model explains the excess returns in the cross-sectional direction. Fama and French (1993) and Fama and French (2015) proxied those factors by the returns on portfolios sorted by different characteristics, and they developed three-factor and five-factor models. After extracting the common movement parts, they treated the intercept as the mispricing *alpha,* which is asset-specific and cannot be explained by those risk factors. Many papers use a similar method to present other factor models, such as the four-factor model of Carhart (1997), the q-factor model of Hou et al. (2015), and the factor zoo by Feng et al. (2017) among others. All of the above papers studied observed factors and did not assign characteristics-based information to either alpha or beta.

Security-specific characteristics, such as capitalization and book to market ratio, are usually documented to explain asset-specific excess returns. Freyberger et al. (2020) analyzed the nonlinear effects of 62 characteristics through Lasso-style regressions. This study concluded that 13 of these characteristics have explanatory power on stock excess returns after selecting by adaptive group Lasso. Characteristics-based information is also exploited to develop arbitrage portfolios by directly parameterizing the portfolio weights as a linear function of characteristics, as in Hjalmarsson and Manchev (2012) and Kim et al. (2019). Empirically, they showed that their portfolio outperformed other baseline competitors.

This paper's contributions are fourfold. Firstly, we build up a more flexible semiparametric characteristics-based asset pricing factor model focusing on the mispricing component. Secondly, we extend previous estimation and testing methods, which can fit the current framework better. Especially, we extend the power-enhanced test of Fan et al. (2015) in a group manner to strengthen the conventional Wald test for mispricing functions. This test can also select the characteristics that contribute to arbitrage portfolios simultaneously. Thirdly, we construct a two-layer clustering structure of mispricing components. Finally, our methods are applied to fifty years of monthly US stock data. We detect distinct clusters of the same characteristics resulting in similar arbitrage returns, forming a "peer group" of arbitrage characteristics. This finding supplements existing portfolio management techniques by implying that the development of arbitrage portfolios through the asset weights determined by the linear mispricing function is improvable.

This class of models has a basic regression specification in Equation 3.1. Consider the panel regression model

$$y_{it} = \alpha_i + \sum_{j=1}^{J} \beta_{ji} f_{jt} + \varepsilon_{it}, \tag{3.1}$$

where y_{it} is the excess return of security i at time t; f_{jt} is the j^{th} risk factor's return at time t; β_{ji} denotes the j^{th} factor loading of asset i; α_i represents the intercept (mispricing) of asset i; and ε_{it} is the mean zero idiosyncratic shock. In terms of factor loadings β_{ji}, Connor and Linton (2007) and Connor et al. (2012) studied a characteristic-beta model, which bridges the beta-coefficients and firm-specific characteristics by specifying each beta as an unknown function of one characteristic. In their model, beta functions and unobservable factors are estimated by the back-fitting iteration. They concluded that those characteristic-beta functions are significant and nonlinear. Their model can be summarized by

$$y_{it} = \sum_{j=1}^{J} g_j(X_{ji}) f_{jt} + \varepsilon_{it}, \tag{3.2}$$

where X_{ji} is the j^{th} observable characteristic of firm i.

They restricted their beta function to be uni-variate and did not consider the components of factor loading functions that cannot be explained by characteristics. To overcome this limitation, Fan et al. (2016b) allowed β_{ji} in Equation 3.2 to have a component explained by observable characteristics as well as an unexplained or stochastic part, written as $\beta_{ji} = g_j(X_i) + u_{ji}$, where u_{ji} is mean independent of X_{ji}. They proposed the Projected Principal Component Analysis (PPCA), which projects stock excess returns onto space spanned by firm-specific characteristics and then applies Principal Component Analysis (PCA) to the projected returns to find the unobservable factors. This method has attractive properties even under the large n and small T setting. However, they did not study the mispricing part (alpha), which is crucial to both asset pricing theories and portfolio management.

In this paper, we work on a semiparametric characteristics-based alpha and beta model, which utilizes a set of security-specific characteristics that are similar to Freyberger et al. (2020). We use unknown multivariate characteristic functions to approximate both α_i and β_{ji} in Equation 3.1. Specifically, we assume α_i and β_{ji} are functions of a large set of asset-specific characteristics as $\alpha_i = h(\boldsymbol{X}_i) + \gamma_i$ and $\beta_{ji} = g_j(\boldsymbol{X}_i) + \lambda_{ij}$[1]. We only specify additive structure of $h(\boldsymbol{X}_i)$ and $g_j(\boldsymbol{X}_i)$, which are further approximated by B-splines sieve. We then estimate

[1] $\boldsymbol{X}_i$ is a vector of a large set of asset-specific characteristics of stock i.

$h(\boldsymbol{X}_i)$, $g_j(\boldsymbol{X}_i)$ and unobservable risk factors f_{jt}. In addition, we design a power-enhanced test and Hierarchical K-mean Clustering for the mispricing function $h(\boldsymbol{X}_i)$ to study the nonlinear behavior of arbitrage characteristics.

Some recent papers such as Kim et al. (2019) and Kelly et al. (2019) analyzed a similar model as ours, which assume that both $h(\boldsymbol{X}_i)$ and $g_j(\boldsymbol{X}_i)$ are *linear functions* . They both included around 40 characteristics in $\boldsymbol{X}_i$. However, they drew different conclusions on the existence of $h(\boldsymbol{X}_i)$. Kim et al. (2019) determined assets weights of arbitrage portfolios using one-year rolling window estimates $\frac{1}{n}\hat{h}(\boldsymbol{X}_i)$. They showed that their arbitrage portfolios returns are statistically and economically significant. However, Kelly et al. (2019) applied instrumented principal component analysis (IPCA) to the entire time span from 1965 to 2014, and concluded no evidence to reject the null hypothesis $H_0 : h(\boldsymbol{X}_i) = \boldsymbol{X}_i^\intercal \boldsymbol{B} = \boldsymbol{0}$ through bootstrap. This dispute spurs the development of a more flexible model and reliable hypothesis tests to investigate the existence and structure of $h(\boldsymbol{X}_i)$. The IPCA, which requires both large n and T to work, was introduced in Kelly et al. (2017). This method does not fit our setting since we apply rolling window analysis with small T. Furthermore, Kelly et al. (2019) restricted the function form of $h(\boldsymbol{X}_i)$ and $g_j(\boldsymbol{X}_i)$ to be time-invariant, which is not consistent with our empirical results under a semiparametric setting. To clarify the differences with the aforementioned research, this paper proposes a semiparametric model, which allows for both nonlinearity and time-variation of $h(\boldsymbol{X}_i)$ and $g_j(\boldsymbol{X}_i)$. Furthermore, we consider a different economic question, namely, the existence and structure of mispricing functions. Our empirical study sheds light on why Kelly et al. (2019) and Kim et al. (2019) drew different conclusions: weak, time-varying and nonlinear characteristics-based mispricing functions only appear in certain rolling windows, which is hard to be detected without rolling window analysis. However, given the presence of some persistent arbitrage characteristics, portfolios developed through mispricing functions can provide arbitrage returns.

The unrestrictive model in this paper brings both opportunities and challenges. According to Huang et al. (2010), the number of B-spline knots must increase in the number of observations to achieve accurate approximation and good asymptotic performance. Therefore, the dimension of B-splines bases coefficients also needs to grow with the sample size. Besides, mispricing functions are treated as anomalies. Under a correctly specified factor model, coefficients of these B-splines bases that are employed to approximate $h(\boldsymbol{X}_i)$ are very likely to be sparse. All of these circumstances make the conventional Wald tests have very low power as discussed in Fan et al. (2015). Therefore, a power-enhanced test should be developed to strengthen the power of Wald tests and to detect the most relevant characteristics among a characteristic zoo included in $h(\boldsymbol{X}_i)$. Kock and Preinerstorfer (2019) illustrated that if

the number of coefficients diverges as the number of observations approaches infinity, the standard Wald test is power enhanceable. Fan et al. (2015) proposed a power-enhanced test by introducing a screening process on all estimated coefficients one by one. They added significant components as a supplement to the standard Wald test. In this paper, we extend Fan et al. (2015) to a group manner to enhance the hypothesis test on a high dimensional additive semiparametric function, $H_0 : h(\boldsymbol{X}_i) = 0$. This method allows all the significant components of $h(\boldsymbol{X}_i)$ to be selected and contribute to the test statistics, with the test power approaching one.

The careful analysis of $h(\boldsymbol{X}_i)$ is theoretically and practically meaningful. Firstly, the presence of $h(\boldsymbol{X}_i)$ is an important component of Arbitrage Pricing Theory (APT) and can contribute to asset pricing theories, namely, linking the mispricing functions with security-related characteristics. Secondly, as in Hjalmarsson and Manchev (2012) and Kim et al. (2019), $h(\boldsymbol{X}_i)$ can be utilized to construct arbitrage portfolios through observed characteristics. However, both research was built upon the condition that $h(\boldsymbol{X}_i)$ is linear over characteristics. If the mispricing function $h(\boldsymbol{X}_i)$ is not monotonic, simply setting portfolio weights to the estimated values of linear-specified $h(\boldsymbol{X}_i)$ can be problematic. In this paper, we show that some characteristics with substantially different values give rise to similar arbitrage returns. The distance of arbitrage returns between two assets i and j is $d_{ij} = |h(\boldsymbol{X}_i) - h(\boldsymbol{X}_j)|$ and the similarity of their characteristics is $\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_2$, where $\|\cdot\|_2$ represents L_2 distance. Inspired by Hoberg and Phillips (2016) and Vogt and Linton (2017), we employ a hierarchical K-means clustering to classify these characteristics within each mispricing return group. We present the dynamic of distinct clusters of the same characteristics leading to similar arbitrage returns, forming a "peer group" of arbitrage characteristics. Therefore, under the semiparametric setting, the asset weighting function should rely on these time-varying and nonlinear peer groups.

The rest of this paper is organized as follows. Section 2 sets out the semiparametric model. Section 3 introduces the assumptions and estimation methods. Section 4 constructs a power-enhanced test for high dimensional additive semiparametric functions. Section 5 employs Hierarchical K-Means Clustering to investigate peer groups of arbitrage characteristics. Section 6 describes the asymptotic properties of our estimates and test statistics. Section 6 simulates data to verify the performance of our methodology. Section 7 presents an empirical study. Finally, Section 8 concludes this paper. Characteristics description tables, proofs, mispricing curves and plots of peer groups are arranged in the Appendix.

3.2 Model Setup

We assume that there are n securities observed over T time periods. We also assume that during a short time window, each security has P time-invariant observed characteristics, such as market capitalization, momentum, and book-to-market ratios. Meanwhile, we may omit heteroskedasticity by assuming that each characteristic shares a certain form of variation within each period for all securities. We suppose that

$$y_{it} = (h(\boldsymbol{X}_i) + \gamma_i) + \sum_{j=1}^{J} (g_j(\boldsymbol{X}_i) + \lambda_{ij}) f_{jt} + \varepsilon_{it}, \tag{3.3}$$

where y_{it} is the monthly excess return of the i^{th} stock at the month t; $\boldsymbol{X}_i$ is a $1 \times P$ vector of P characteristics of stock i during time periods $t = 1, \dots T$. T is a small and fixed time block. In practice, most characteristics are updated annually. Thus, we assume $\boldsymbol{X}_i$ is time-invariant in one-year time window. $h(\boldsymbol{X}_i)$ is an unknown mispricing function explained by a large set of characteristics whereas γ_i is the random intercept of the mispricing part that cannot be explained by characteristics. Similarly, we have characteristics-beta function $g_j(\cdot)$ to explain the j^{th} factor loadings and the unexplained stochastic part of the loading is λ_{ij}. λ_{ij} is orthogonal to the $g_j(\cdot)$ function. f_{jt} is the realization of the j^{th} risk factor at time t. Finally, ε_{it} is homoskedastic zero-mean idiosyncratic residual of the i^{th} stock at time t. Random variables γ_i and λ_{ij} are used to generalize our settings and not to be estimated. They will be treated as noise in the identification assumptions.

To avoid the curse of dimensionality, we impose additive forms on both $h(\cdot)$ and $g_j(\cdot)$ functions: $h(\boldsymbol{X}_i) = \sum_{p=1}^{P} \mu_p(X_{ip})$ and $g_j(\boldsymbol{X}_i) = \sum_{p=1}^{P} \theta_{jp}(X_{ip})$, where $\mu_p(X_{ip})$ and $\theta_{jp}(X_{ip})$ are uni-variate unknown functions of the p^{th} characteristic X_p. We rewrite the model as:

$$y_{it} = (\sum_{p=1}^{P} \mu_p(X_{ip}) + \gamma_i) + \sum_{j=1}^{J} (\sum_{p=1}^{P} \theta_{jp}(X_{ip}) + \lambda_{ij}) f_{jt} + \varepsilon_{it}, \tag{3.4}$$

Assumption 10. *We suppose that:*

$$E(\varepsilon_{it}|\boldsymbol{X}, f_{jt}) = 0,$$

$$E(h(\boldsymbol{X}_i)) = E(g_j(\boldsymbol{X}_i)) = 0, \textit{ for } j = 1, 2, \dots, J$$

$$E(\gamma_i|\boldsymbol{X}) = E(\gamma_i),$$

$$E(\lambda_{ij}|\boldsymbol{X}) = E(\lambda_{ij}), \textit{ for } j = 1, 2, \dots, J$$

$$E(h(\boldsymbol{X}_i)g_j(\boldsymbol{X}_i)) = \mathbf{0}, \textit{ for } j = 1,2,\ldots,J$$

Similar to Connor et al. (2012) and Fan et al. (2016b), Assumption 10 above is to standardize model settings, including the zero mean assumption for factor loadings and mispricing functions for identification purposes. We also impose orthogonality between mispricing and factor loading parts for the identification reason. This is because the variation of risk factors can be absorbed into the mispricing part if it is not orthogonal to the factor loadings. More discussions can be found in Connor et al. (2012).

3.3 Estimation

In this section we discuss the approximation of unknown uni-variate functions and our estimation methods for model Equation 3.3. In the semiparametric setting, we apply the Projected-PCA following Fan et al. (2016b) to work on the common factors and characteristics-beta directly. Next, we project the residuals onto the characteristics-based alpha space that is orthogonal to the systematic part. The second step is similar to equality-constrained OLS.

3.3.1 B-splines Approximation

We use B-splines sieve to approximate unknown functions $\theta(\cdot)$ and $\mu(\cdot)$ in Equation 3.4. Similar to Huang et al. (2010) and Chen and Pouzo (2012), we have the following procedures. Firstly, suppose that the p^{th} covariate X_p is in the interval $[D_0, D]$, where D_0 and D are finite numbers with $D_0 < D$. Let $\boldsymbol{D} = \{\underbrace{D_0, D_0, \ldots, D_0}_{l} < d_1 < d_2 < \cdots < d_{m_n} < \underbrace{D, D, \ldots, D}_{l}\}$ be a simple knot sequence on the interval $[D_0, D]$. Here, $m_n = \lfloor n^v \rceil$ ($\lfloor \cdot \rceil$ gives nearest integer) is a positive integer of the number of internal knots, which is a function of security size n in period t with $0 < v < 0.5$. l is the degree of those bases. Therefore, we have $H_n = l + m_n$ bases in total, which will diverge as $n \to \infty$. Following this setting, a set of B-splines can be built for the space $\Omega_n[\boldsymbol{D}]$.

Secondly, for the p^{th} characteristic X_p, there is a set of H_n orthogonal bases $\{\phi_{1p}(X_p), \ldots, \phi_{H_n p}(X_p)\}$. Those uni-variate unknown functions can be approximated as linear combinations of these bases as $\mu_p(X_p) = \sum_{q=1}^{H_n} \alpha_q \phi_{qp}(X_p) + R_p^{\mu}(X_p)$ and $\theta_p(X_p) = \sum_{q=1}^{H_n} \beta_{jq} \phi_{qp}(X_p) + R_p^{\theta}(X_p)$, where $R_p^{\mu}(X_p)$ and $R_p^{\theta}(X_p)$ are approximation errors. It is not necessary to use the same bases for both unknown functions and the representation here is for notational simplicity only. Therefore, the model Equation 3.4 can be written as:

$$y_{it} = \sum_{p=1}^{P}(\sum_{q=1}^{H_n} \alpha_{pq}\phi_{pq}(X_{ip}) + R_p^{\mu}(X_p)) + \gamma_i + \sum_{j=1}^{J}(\sum_{p=1}^{P}(\sum_{q=1}^{H_n} \beta_{jpq}\phi_{pq}(X_{ip}) + R_p^{\theta}(X_p)) + \lambda_{ij})f_{jt} + \varepsilon_{it}$$

For each $i = 1,2,\ldots,n$, $p = 1,2,\ldots,P$ and $t = 1,2,\ldots,T$, we have:

$$\mathbf{1}_T = (1,\ldots,1)^{\intercal} \in \mathbb{R}^T,$$

$$\beta_j = (\beta_{1,j1},\ldots,\beta_{H_n,j1},\ldots,\beta_{1,jP},\ldots,\beta_{H_n,jP})^{\intercal} \in \mathbb{R}^{H_nP},$$

$$\boldsymbol{B} = (\beta_1,\ldots,\beta_J),$$

$$\boldsymbol{A} = (\alpha_{11},\ldots,\alpha_{1H_n},\ldots,\alpha_{P1},\ldots,\alpha_{PH_n})^{\intercal} \in \mathbb{R}^{H_nP},$$

$$\boldsymbol{\Phi}(\boldsymbol{X}) = \begin{bmatrix} \phi_{1,11}(X_{11}) & \cdots & \phi_{1,1H_n}(X_{11}) & \cdots & \phi_{1,P1}(X_{1P}) & \ldots & \phi_{1,PH_n}(X_{1P}) \\ \phi_{2,11}(X_{21}) & \cdots & \phi_{2,1H_n}(X_{21}) & \cdots & \phi_{2,P1}(X_{2P}) & \ldots & \phi_{2,PH_n}(X_{2P}) \\ \vdots & \vdots & \vdots & \ddots & \vdots & & \\ \phi_{n,11}(X_{n1}) & \cdots & \phi_{n,1H_n}(X_{n1}) & \cdots & \phi_{n,P1}(X_{nP}) & \ldots & \phi_{n,PH_n}(X_{nP}) \end{bmatrix},$$

where $\phi_{i,ph}(X_{ip})$ is the h^{th} basis of the p^{th} characteristic of asset i at time t. Therefore, the original model

$$\boldsymbol{Y} = (h(\boldsymbol{X}) + \boldsymbol{\Gamma})\mathbf{1}_T^{\intercal} + (\boldsymbol{G}(\boldsymbol{X}) + \boldsymbol{\Lambda})\boldsymbol{F}^{\intercal} + \boldsymbol{U},$$

can be represented by B-splines sieve as:

$$\boldsymbol{Y} = (\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A} + \boldsymbol{\Gamma} + \boldsymbol{R}^{\mu}(\boldsymbol{X}))\mathbf{1}_T^{\intercal} + (\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B} + \boldsymbol{\Lambda} + \boldsymbol{R}^{\theta}(\boldsymbol{X}))\boldsymbol{F}^{\intercal} + \boldsymbol{U}, \tag{3.5}$$

$\boldsymbol{Y}$ is the $n \times T$ matrix of y_{it}; $\boldsymbol{\Phi}(\boldsymbol{X})$ is the $n \times PH_n$ matrix of B-splines bases; $\boldsymbol{A}$ is a $PH_n \times 1$ matrix of mispricing coefficients; $\boldsymbol{R}^{\mu}(\boldsymbol{X})$ is a $n \times 1$ matrix of approximation errors; $\boldsymbol{B}$ is a $PH_n \times J$ matrix factor loadings' coefficients; $\boldsymbol{R}^{\theta}(\boldsymbol{X})$ is a $n \times J$ matrix of approximation errors. We have $R_p^{\mu}(X_p) \to^p 0$ and $R_p^{\theta}(X_p) \to^p 0$, as $n \to \infty$ as in Huang et al. (2010). Therefore, we omit the approximation errors for simplicity henceforth. $\boldsymbol{F}$ is the $T \times J$ matrix of f_{tj} and $\boldsymbol{U}$ is a $n \times T$ matrix of ε_{it}. $h(\boldsymbol{X})$ is a $n \times 1$ vector of characteristics-based mispricing component; $\boldsymbol{G}(\boldsymbol{X})$ is a $n \times J$ vector of characteristics-based factor loadings; $\mathbf{1}_T$ is a $T \times 1$ vector of 1. The rest are defined the same as Equation 3.4.

We define a projection matrix as:

$$\boldsymbol{P} = \boldsymbol{\Phi}(\boldsymbol{X})(\boldsymbol{\Phi}(\boldsymbol{X})^{\intercal}\boldsymbol{\Phi}(\boldsymbol{X}))^{-1}\boldsymbol{\Phi}(\boldsymbol{X})^{\intercal}.$$

The remaining goals of this paper are to estimate both $h(\boldsymbol{X})$ and $\boldsymbol{G}(\boldsymbol{X})$ consistently and conduct a power-enhanced test on the hypothesis $H_0 : h(\boldsymbol{X}) = \boldsymbol{0}$, i.e., to check the existence of mispricing functions under semiparametric settings. Finally, we cluster peer groups of arbitrage characteristics.

3.3.2 Two-Step Projected-PCA

In this section, we combine and extend Projected-PCA by Fan et al. (2016b) and equality constrained least squares similar to Kim et al. (2019) to estimate the model. To facilitate the estimation, we define a $T \times T$ time series demeaning matrix $\boldsymbol{D}_T = \boldsymbol{I}_T - \frac{1}{T}\boldsymbol{1}_T\boldsymbol{1}_T^\intercal$.[2] Next, we demean the equation above on both sides. Therefore we have

$$\boldsymbol{Y}\boldsymbol{D}_T = \tilde{\boldsymbol{Y}} = (\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B} + \boldsymbol{\Lambda})\boldsymbol{F}^\intercal\boldsymbol{D}_T + \boldsymbol{U}\boldsymbol{D}_T.$$

Mispricing terms disappear since they are time-invariant by $(\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A} + \boldsymbol{\Gamma})\boldsymbol{1}_T^\intercal\boldsymbol{D}_T = \boldsymbol{0}$. This helps us to work on the systematic part later. Henceforth, we use $\boldsymbol{F}$ to represent the time-demeaned factor matrix.

Our procedures are designed to estimate factor loadings $\boldsymbol{G}(\boldsymbol{X})$, time-demeaned unobserved factors $\boldsymbol{F}$ and mispricing coefficients $\boldsymbol{A}$ in sequence.

Under Assumption 10, we have the following estimation procedures:

1 Projecting $\tilde{\boldsymbol{Y}}$ onto the spline space spanned by $\{\boldsymbol{X}_{ip}\}_{i\leqslant n, p\leqslant P}$ through a $n \times n$ projection matrix $\boldsymbol{P}$ with $\boldsymbol{P} = \boldsymbol{\Phi}(\boldsymbol{X})(\boldsymbol{\Phi}(\boldsymbol{X})^\intercal\boldsymbol{\Phi}(\boldsymbol{X}))^{-1}\boldsymbol{\Phi}(\boldsymbol{X})^\intercal$. We then collect the projected data $\hat{\boldsymbol{Y}} = \boldsymbol{\Phi}(\boldsymbol{X})(\boldsymbol{\Phi}(\boldsymbol{X})^\intercal\boldsymbol{\Phi}(\boldsymbol{X}))^{-1}\boldsymbol{\Phi}(\boldsymbol{X})^\intercal\tilde{\boldsymbol{Y}}$.

2 Applying the Principal Component Analysis to the projected data $\hat{\boldsymbol{Y}}^\intercal\hat{\boldsymbol{Y}}$. This allows us to work directly on the sample covariance of $\boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}^\intercal$, under the condition $E(g_j(\boldsymbol{X}_i)\varepsilon_{it}) = E(g_j(\boldsymbol{X}_i)\lambda_{ij}) = 0$.

3 Estimating $\hat{\boldsymbol{F}}$ as the eigenvectors corresponding to the first J (assumed given) eigenvalues of the $T \times T$ matrix $\frac{1}{n}\hat{\boldsymbol{Y}}^\intercal\hat{\boldsymbol{Y}}$ (covariance of projected $\hat{\boldsymbol{Y}}$).

The method above substantially improves estimation accuracy and facilitates theoretical analysis even under the large n and small T . Small T is preferable in our model setting as we use one-year rolling windows analysis in both simulation and empirical studies, and large n is required for asymptotic analysis.

[2] $\boldsymbol{I}_T$ is a $T \times T$ identity matrix, and $\boldsymbol{1}_T$ is a $T \times 1$ matrix of 1.

Factor loadings $\hat{\boldsymbol{G}}(\boldsymbol{X})$ are estimated as:

$$\hat{\boldsymbol{G}}(\boldsymbol{X}) = \hat{\boldsymbol{Y}}\hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}^{\intercal}\hat{\boldsymbol{F}})^{-1}$$

In the next step, we estimate the coefficients of the mispricing bases.

4 The estimator of $\boldsymbol{A}$ is

$$\hat{\boldsymbol{A}} = \arg\min_{\boldsymbol{A}} \operatorname{vec}(\boldsymbol{Y} - \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}\mathbf{1}_T^{\intercal} - \hat{\boldsymbol{G}}(\boldsymbol{X})\hat{\boldsymbol{F}}^{\intercal})^{\intercal}\operatorname{vec}(\boldsymbol{Y} - \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}\mathbf{1}_T^{\intercal} - \hat{\boldsymbol{G}}(\boldsymbol{X})\hat{\boldsymbol{F}}^{\intercal}), \quad (3.6)$$

subject to $\hat{\boldsymbol{G}}(\boldsymbol{X})^{\intercal}\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A} = \mathbf{0}_J$.

Let a $PH_n \times 1$ vector $\hat{\boldsymbol{A}}$ be a closed-form solution:

$$\hat{\boldsymbol{A}} = \boldsymbol{Q}\tilde{\boldsymbol{A}},$$

where

$$\boldsymbol{Q} = \boldsymbol{I} - (\boldsymbol{\Phi}(\boldsymbol{X})^{\intercal}\boldsymbol{\Phi}(\boldsymbol{X}))^{-1}\boldsymbol{\Phi}(\boldsymbol{X})^{\intercal}\hat{\boldsymbol{G}}(\boldsymbol{X})(\hat{\boldsymbol{G}}(\boldsymbol{X})^{\intercal}\hat{\boldsymbol{G}}(\boldsymbol{X}))^{-1}\hat{\boldsymbol{G}}(\boldsymbol{X})^{\intercal}\boldsymbol{\Phi}(\boldsymbol{X}),$$

$$\tilde{\boldsymbol{A}} = \frac{1}{T}(\boldsymbol{\Phi}(\boldsymbol{X})^{\intercal}\boldsymbol{\Phi}(\boldsymbol{X}))^{-1}\boldsymbol{\Phi}(\boldsymbol{X})^{\intercal}(\boldsymbol{Y} - \hat{\boldsymbol{G}}(\boldsymbol{X})\hat{\boldsymbol{F}}^{\intercal})\mathbf{1}_T,$$

given $\boldsymbol{P}\hat{\boldsymbol{G}}(\boldsymbol{X}) = \hat{\boldsymbol{G}}(\boldsymbol{X})$.

As in Assumption 10, the $h(\boldsymbol{X})$ is orthogonal to the characteristics-based loadings $\boldsymbol{G}(\boldsymbol{X})$.

5 We also estimate the covariance matrix of $\hat{\boldsymbol{A}}$, i.e., $\boldsymbol{\Sigma}$, by extending the methods of Liew (1976). This can facilitate theoretical analysis in the next section. According to Liew (1976), $\hat{\boldsymbol{A}}$ is the equality constrained least-square estimator, which has the covariance matrix as (under $n \leqslant T$ and covariance shrinkage as in Ledoit et al. (2012) and Fan et al. (2013) among others.):

$$\hat{\boldsymbol{\Sigma}} = \boldsymbol{Q}\hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{A}}}\boldsymbol{Q}^{\intercal},$$

where:

$$\hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{A}}} = (\boldsymbol{\Phi}(\boldsymbol{X})^{\intercal}\boldsymbol{\Phi}(\boldsymbol{X}))^{-1}\boldsymbol{\Phi}(\boldsymbol{X})^{\intercal}\begin{bmatrix}\hat{\sigma}_1^2 & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \hat{\sigma}_n^2\end{bmatrix}\boldsymbol{\Phi}(\boldsymbol{X})(\boldsymbol{\Phi}(\boldsymbol{X})^{\intercal}\boldsymbol{\Phi}(\boldsymbol{X}))^{-1},$$

$$\hat{\sigma}_i^2 = \frac{\sum_1^T \hat{e}_{it}^2}{T-1},$$

where $\sum_1^T \hat{e}_{it}^2 = \sum_1^T (y_{it} - \bar{y}_i - \sum_{p=1}^P \sum_{q=1}^{H_n} \hat{\alpha}_{pq}\phi_{pq}(x_{ip}) - \sum_{j=1}^J (\sum_{p=1}^P \sum_{q=1}^{H_n} \hat{\beta}_{jpq}\phi_{pq}(x_{ip}))\hat{f}_{jt})^2$. Heteroskedasticity is caused by γ_i.

3.4 Power-enhanced Tests

There are considerable discussions about the mispricing phenomenon under factor models, while the existence of mispricing functions remains controversial. Namely, whether there are relevant covariates explaining remaining excess returns after subtracting co-movements components captured by risk factors. Recently, Kim et al. (2019) found the characteristics arbitrage opportunities by estimating a linear characteristic mispricing function without providing theoretical results. However, Kelly et al. (2019) conducted a conventional Wald hypothesis test on the similar mispricing function using bootstrap, concluding that there is no evidence to reject the null hypothesis $H_0 : h(\boldsymbol{X}) = \mathbf{0}$. Additionally, they applied the bootstrap method to estimate the covariance matrix $\boldsymbol{\Sigma}$, which caused potential problems for theoretical analysis. Moreover, according to Fan et al. (2015), their test results may have relatively low power when the true coefficient vector of linear mispricing function $\boldsymbol{A}$ has a sparse structure.

Both studies adopt a parametric framework, which relies on the strong assumption of linearity. However, this assumption is not consistent with Connor et al. (2012), which showed that both characteristic-beta and mispricing functions are very likely to be nonlinear. Therefore, we propose a semiparametric model to accommodate the nonlinearity to a great extent.

But semiparametric framework leads to additional challenges for inference. On the one hand, as mentioned above, the number of coefficients of mispricing B-splines diverges as $n \to \infty$, which implies that the power of the standard Wald test can be quite low, (see Fan et al. (2015)). On the other hand, according to other research like Fama and French (1993) and Fama and French (2015), mispricing terms can be regarded as anomalies. This means that in our model setting, the true mispricing coefficient vector $\boldsymbol{A}$ can be high-dimensional but sparse, reducing the power of the conventional Wald test further.

As in Kock and Preinerstorfer (2019), conventional hypothesis tests under these circumstances are power enhanceable. The power-enhanced Wald test in this paper is an extension of Fan et al. (2015) to a group manner, namely, the hypothesis test under high-dimensional additive semiparametric settings. The proposed test is power strengthened when the dimension of the coefficients of the additive regression $\boldsymbol{A}$ is diverging as $n \to \infty$ without size distortion.

Meanwhile, this test is robust to sparse alternatives. On top of that, the proposed test can select the most important components from sparse additive functions. Finally, the proposed method can also be applied when the number of characteristics is diverging, i.e., $P \to \infty$.

We construct a new test:

$$H_0 : h(\boldsymbol{X}) = \mathbf{0}, \qquad H_1 : h(\boldsymbol{X}) \neq \mathbf{0},$$

equivalently,

$$H_0 : \boldsymbol{A} = \mathbf{0}, \qquad H_1 : \boldsymbol{A} \in \mathscr{A},$$

where $\mathscr{A} \subset \mathbb{R}^{PH_n} \backslash \mathbf{0}$.

Here, we have:

$$S_1 = \frac{\hat{\boldsymbol{A}}^\intercal \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{A}} - PH_n}{\sqrt{2PH_n}},$$

where S_1 is the "original" Wald test statistics; P is the number of characteristics; PH_n is the total number of B-spline bases, and $\hat{\boldsymbol{A}} \in \mathbb{R}^{PH_n}$. The value of H_n is a function of asset number n, therefore, $H_n \to \infty$ as $n \to \infty$. Under H_0, S_1 has a nondegenerate limiting distribution F as $n \to \infty$. Given the significance level q, $q \in (0,1)$ as well as the critical value F_q:

$$S_1 | H_0 \to^d F,$$

$$\lim_{n\to\infty} \Pr(S_1 > F_q | H_0) = q.$$

Pesaran and Yamagata (2012) showed that:

$$S_1 | H_0 \to^d \mathscr{N}(0,1),$$

under regularity conditions.

Potentially, sparse and diverging PH_n means that it is plausible to add a power-enhanced component to S_1, which can improve the power of the hypothesis test without any size distortions.

Therefore, we can construct an extra screening component S_0 as:

$$S_0 = H_n \sum_{p=1}^{P} \boldsymbol{I}\Big(\sum_{h=1}^{H_n} |\hat{\alpha}_{ph}| / \hat{\sigma}_{ph} \geqslant \eta_n\Big),$$

where $\hat{\sigma}_{ph}$ is the square-root of the ph^{th} entry of the diagonal elements of $\hat{\boldsymbol{\Sigma}}$. $\boldsymbol{I}(\cdot)$ is an indicator for the screening process while η_n is a data-driven threshold value to avoid potential size-distortion.

Here we discuss the choice of η_n. By construction and the assumption of independent characteristics, we assume that all B-splines bases are orthogonal. Our goal is to bound the maximum of those standardized coefficients under the null hypothesis.

Define $Z = \max_{\{1\leqslant p\leqslant P, 1\leqslant h\leqslant H_n\}} \{|\hat{\alpha}_{ph}|/\hat{\sigma}_{ph}\}$. We have

$$\hat{\alpha}_{ph}/\hat{\sigma}_{ph}|\boldsymbol{H}_0 \to^d N(0,1),$$

$$E(Z|\boldsymbol{H}_0) = \sqrt{2\log PH_n}.$$

After grouping the coefficients of bases used to approximate the unknown function of each characteristic, let $R = \max(\sum_{h=1}^{H_n} |\hat{\alpha}_{1h}|/\hat{\sigma}_{1h}, \ldots, \sum_{h=1}^{H_n} |\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \ldots, \sum_{h=1}^{H_n} |\hat{\alpha}_{Ph}|/\hat{\sigma}_{Ph})$. Following this, we set the threshold as $\eta_n = H_n\sqrt{2\log(PH_n)}$, where $H_n = l + n^v$. As H_n is a diverging sequence, it can control the influence of the group size properly. Meanwhile, η_n also diverges so that η_n is a conservative threshold value used to avoid potential size distortion.

Apart from strengthening the power of conventional hypothesis test, $\boldsymbol{I}(\cdot)$ is a screening term that can select the most relevant characteristics at the same time.

We then define the arbitrage characteristics set, which includes the characteristics that have the strong explanation power for mispricing functions:

$$\hat{\mathscr{M}} = \{\boldsymbol{X}_p \in \boldsymbol{X} : \sum_{h=1}^{H_n} |\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \geqslant \eta_n, \quad p = 1,2,\ldots,P\},$$

and M is the cardinality of the set containing mispricing characteristics. When the set $\hat{\mathscr{M}}$ is relatively small, conventional tests are likely to suffer the lower power problem. The added S_0 strengthens the power of the test and drives the power to one since S_0 is slowly diverging.

Therefore, our new test statistic is $S = S_0 + S_1$, and asymptotic properties of S will be discussed later.

To conclude, the advantages of $S = S_0 + S_1$ are:

1 The power of the hypothesis test on $H_0 : h(\boldsymbol{X}) = \mathbf{0}$ is mainly enhanced without size distortions.

2 We can find specific characteristics which cause the mispricing by this screening mechanism.

As designed, S_0 satisfies all three properties of Fan et al. (2015), as $n \to \infty$:

1 S_0 is non-negative, $\Pr(S_0 \geqslant 0) = 1$

2 S_0 does not cause size distortion: under H_0, $\Pr(S_0 = 0 \mid H_0) \to 1$

3 S_0 enhances test power. Under H_1, S_0 diverges quickly in probability given the well chosen η_n.

Based on properties of S_0, we have three properties of S listed:

1 No size-distortion $\limsup_{n\to\infty} \Pr(S > F_q | H_0) = q$

2 $\Pr(S > F_q | H_1) \geqslant \Pr(S_1 > F_q | H_1)$. Hence, the power of S is at least as large as that of S_1.

3 $\Pr(S > F_q | \hat{\mathscr{M}} \neq \emptyset) \to 1$ when S_0 diverges. This happens, especially, when the true form of $\boldsymbol{A}$ has a sparse structure.

3.5 Hierarchical K-Means Clustering

This section introduces a Hierarchical K-means Clustering method to find peer groups of arbitrage characteristics based on their arbitrage returns. We ask whether distinct groups of the same arbitrage characteristics, according to their similarity measured by $\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_2$, may result in similar characteristic-based arbitrage returns in each rolling block, which is an implication for the non-monotonic mispricing function, and forms a "peer group" of arbitrage characteristics. Because traditional arbitrage portfolios as in Kim et al. (2019) and Hjalmarsson and Manchev (2012) rely on the linearity of characteristics-based mispricing components to work, our clustering results can show whether this method is still applicable under more flexible semiparametric model. If there are persistent peer groups in arbitrage returns, investors should consider to long the assets in the peer group with the highest

arbitrage returns while short the assets in the peer group with the lowest arbitrage returns to form an arbitrage portfolio.

Introduction of K-means clustering can be found in Cox (1957) and Fisher (1958).

After the screening process in section 3.4, we obtain the relevant components of mispricing function $h(\boldsymbol{X})$, which is estimated as

$$\hat{\mathscr{M}} = \{\boldsymbol{X}_p \in \boldsymbol{X} : \sum_{h=1}^{H_n} |\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \geqslant \eta_n, \quad p = 1,2,\ldots,P\}.$$

We define an $n \times M$ matrix $\boldsymbol{M}$ of arbitrage characteristics at time window t as :

$$\boldsymbol{M} = \{\boldsymbol{X}_1,\ldots,\boldsymbol{X}_m,\ldots,\boldsymbol{X}_M\}, \text{ where } \boldsymbol{X}_m \in \hat{\mathscr{M}}.$$

Note that these characteristics are time-invariant in each rolling window. We also set characteristics-based arbitrage returns of asset i in month t as:

$$\ddot{y}_{it} = \boldsymbol{\phi}(\boldsymbol{M}_i)\hat{\boldsymbol{A}}_M,$$

where $\boldsymbol{\phi}(\boldsymbol{M}_i)$ and $\hat{\boldsymbol{A}}_M$ are the corresponding parts of matrix $\boldsymbol{\Phi}(\boldsymbol{X}_i)$ and vector $\hat{\boldsymbol{A}}$. For each rolling window, we classify all n assets through a 2-layer K-means clustering. At the first layer, we cluster these assets into K groups according to the similarity of their characteristics-based arbitrage returns $\ddot{y}_{it}$. At the second layer, we divide R_j subgroups within the j^{th} group from the first layer by the similarity of their arbitrage characteristics, where $j = 1,2,\ldots,K$. Finally, the peer groups of arbitrage characteristics can be attained. We repeat this method for all rolling blocks to investigate dynamic patterns of these peer groups. These clusterings will provide illustrative evidence of linear/nonlinear and time-invariant/time-varying structure of mispricing function $h(\boldsymbol{X})$.

We give the classification procedures of both layers. We define Δ_{ij} as the difference between characteristics-based arbitrage returns of $\ddot{y}_{it}$ and $\ddot{y}_{jt}$, as well as Υ_{ij} as the difference between arbitrage characteristics:

$$\Delta_{ij} = \ddot{y}_{it} - \ddot{y}_{jt}, \text{ where } i \neq j, \quad i,j = 1,2,\ldots,n.$$

$$\Upsilon_{ij} = \|\boldsymbol{M}_i - \boldsymbol{M}_j\|_2, \text{ where } i \neq j, \quad i,j = 1,2,\ldots,n,$$

$\boldsymbol{M}_i$ represents the i^{th} row of $\boldsymbol{M}$. We set two tolerance thresholds ψ_y and ψ_x, which are used to control the biggest difference within each group of both layers separately. To accelerate the convergence of the K-means Clustering, we first apply a first difference process, which is introduced below, to obtain centroids as in Vogt and Linton (2017).

For the first layer, we have first difference process:

1. **First difference**: We randomly pick i^{th} asset and then we calculate Δ_{ij} with other assets $j = 1,2,\ldots,n$. Thus we obtain $\Delta_{i(1)} \ldots \Delta_{i(n)}$, with n being the total individuals for classification. Without loss of generality, we assume $\Delta_{i(1)} = \min\{\Delta_{i(1)},\ldots,\Delta_{i(n)}\}$, and $\Delta_{i(n)} = \max\{\Delta_{i(1)},\ldots,\Delta_{i(n)}\}$.

2. **Ordering**: We rank the values obtained in Step 1 as follows:

$$\begin{aligned}\Delta_{i(1)} \leqslant \ldots \leqslant \Delta_{i(j_1-1)} &< \Delta_{i(j_1)} \leqslant \ldots \leqslant \Delta_{i(j_2-1)} \\ &< \Delta_{i(j_2)} \leqslant \ldots \leqslant \Delta_{i(j_3-1)} \\ &\vdots \\ &< \Delta_{i(j_{K-1})} \leqslant \ldots \leqslant \Delta_{i(n)}.\end{aligned}$$

   We use the strict inequality mark to show large jumps of "first difference", all of which are larger than ψ_y , while the weak inequality means that the distance calculated is not larger than ψ_y. We identify $K-1$ jumps that are larger than ψ_y above. Thus, the initial classification is achieved, and we have a total of K groups with $j_1 - 1$ members in the first group $\mathscr{C}_1$, $j_2 - j_1$ members in the second group $\mathscr{C}_2$, ..., and $n - j_{K-1} + 1$ members in the final group $\mathscr{C}_K$.

In terms of the second layer, for the assets in the k^{th} group $\mathscr{C}_k$, we use the same method to further divide them into r subgroups as $\mathscr{R}_{1k}, \mathscr{R}_{2k}, \ldots, \mathscr{R}_{rk}$. Within each subgroup, we have:

$$\Upsilon_{ab} = \|\boldsymbol{M}_a - \boldsymbol{M}_b\|_2 \leqslant \psi_x, \text{ where } a,b \in \mathscr{R}_{ik}, i = 1,2,\ldots,r, \text{ and } k = 1,2,\ldots,K.$$

The K-means algorithm is:

1. Step 1: Determine the starting mean values for each group $\hat{\ddot{c}}_1^{[0]},\ldots,\hat{\ddot{c}}_K^{[0]}$ and calculate the distances $\hat{D}_k(i) = \Delta(\ddot{y}_{it}, \hat{\ddot{c}}_k^{[0]}) = |\ddot{y}_{it} - \hat{\ddot{c}}_k^{[0]}|$ for each i and k. Define the partition $\{\mathscr{C}_1^{[0]},\ldots,\mathscr{C}_K^{[0]}\}$ by assigning the i^{th} individual to the k-th group $\mathscr{C}_k^{[0]}$ if $\hat{D}_k(i) = \min_{1\leqslant k'\leqslant K} \hat{D}_{k'}(i)$.

2. Step l: Let $\{\mathscr{C}_1^{[l-1]},\ldots,\mathscr{C}_K^{[l-1]}\}$ be the partition of $\{1,\ldots,n\}$ from the latest iteration step. Calculate mean functions

$$\hat{\ddot{c}}_k^{[l]} = \frac{1}{|\mathscr{C}_k^{[l-1]}|}\sum_{i\in\mathscr{C}_k^{[l-1]}} \ddot{y}_{it} \qquad \text{for } 1\leqslant k\leqslant K$$

   And then we calculate $\Delta(\ddot{y}_{it},\hat{\ddot{c}}_k^{[l]}) = |\ddot{y}_{it}-\hat{\ddot{c}}_k^{[l]}|$ for each i and k. Define the partition $\{\mathscr{C}_1^{[l]},\ldots,\mathscr{C}_K^{[l]}\}$ by assigning the i^{th} individual to the k-th group $\mathscr{C}_k^{[l]}$ if $\hat{D}_k(i) = \min_{1\leqslant k'\leqslant K}\hat{D}_{k'}(i)$.

3. Iterate the above steps until the partition $\{\mathscr{C}_1^{[w]},\ldots,\mathscr{C}_K^{[w]}\}$ does not change anymore.

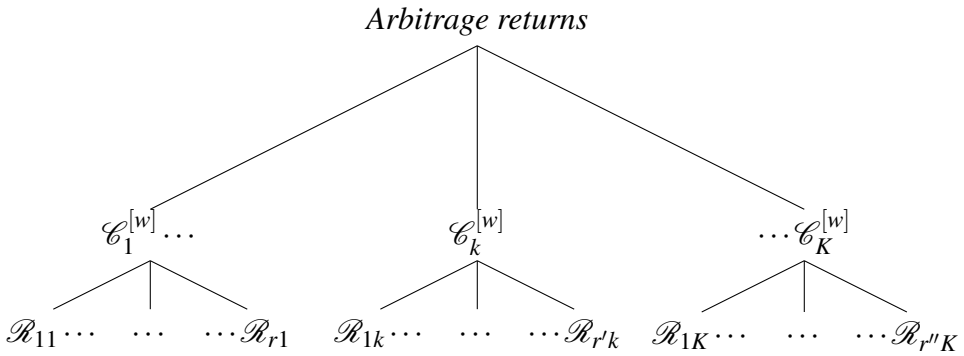
To accelerate the convergence of K-means algorithm, at the step 1, results of first difference are used. As we have already obtained our initial grouping $\{\mathscr{C}_1,\ldots,\mathscr{C}_K\}$, therefore starting values for the Step 1 is:

$$\hat{\ddot{c}}_k^{[0]} = \frac{1}{|\mathscr{C}_k|}\sum_{i\in\mathscr{C}_k} \ddot{y}_{it} \qquad \text{for } 1\leqslant k\leqslant K,$$

where $|\mathscr{C}_k|$ is the cardinality of the group $\mathscr{C}_k$.

The consistency and other theoretical results of the above procedures can be found in Pollard (1981), Pollard et al. (1982), Sun et al. (2012) and Vogt and Linton (2017).

For the second layer, we repeat the procedures within each group $\mathscr{C}_k^{[w]}$ with respect to Υ_{ab}, and the structure of characteristics-based arbitrage returns is:



The first layer is the structure of characteristics-based arbitrage returns, while the second layer gives peer groups of characteristics that can provide similar characteristics-based arbitrage returns.

The number of clusterings is determined by threshold values ψ_y and ψ_x directly. ψ_y and ψ_x are chosen by the trade-off between the number of clusterings and total within-group sum of squares.

3.6 Asymptotic Properties

This section discusses assumptions and properties of estimates and power enhanced statistics S.

Definition 2. We define $\boldsymbol{A} \to^P \boldsymbol{B}$ as $n \to \infty$ of two $n \times m$ matrix $\boldsymbol{A}$ and $\boldsymbol{B}$ with fixed p when $\frac{1}{n}(\boldsymbol{A}-\boldsymbol{B})^\intercal(\boldsymbol{A}-\boldsymbol{B}) \to^P \boldsymbol{0}_{m\times m}$ as $n \to \infty$.

3.6.1 Consistency Assumptions

Assumption 11. *As $n \to \infty$, we have:*

$$\frac{1}{n}\boldsymbol{Y}^\intercal\boldsymbol{Y} \to_P \boldsymbol{M}_Y,$$

$$\boldsymbol{F}^\intercal\boldsymbol{F} = \boldsymbol{I}_J,$$

where $\boldsymbol{M}_Y$ is a positive definite matrix, and $\boldsymbol{I}_J$ is a $J \times J$ identity matrix.

We define $\lambda_{min}(M)$ and $\lambda_{max}(M)$ as the largest and the smallest eigenvalues of matrix M, respectively. Additionally, we define C_{min} and C_{max} to be positive constants such that:

$$C_{min} \leqslant \lambda_{min}(\frac{1}{n}\boldsymbol{\Phi}^\intercal(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})) < \lambda_{max}(\frac{1}{n}\boldsymbol{\Phi}^\intercal(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})) \leqslant C_{max}$$

as $n \to \infty$.

We impose these restrictions to avoid non-invertibility of stock returns, characteristics, and rotation indeterminacy.

Assumption 12.

$$\frac{1}{n}\boldsymbol{G}(\boldsymbol{X})^\intercal\boldsymbol{G}(\boldsymbol{X}) \to_P \begin{bmatrix} d_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & d_J \end{bmatrix},$$

as $n \to \infty$, where d_j are distinct and positive entries.

Both Assumption 11 and 12 are similar to those in Fan et al. (2016b), which are used to separately identify risk factors and factor loadings. Given the orthogonal bases of B-splines and uncorrelated or weakly correlated characteristics, Assumption 12 is mild.

Assumption 13. *K_{min} and K_{max} are positive constants such that:*

$$K_{min} \leqslant \lambda_{min}(\frac{1}{n}\boldsymbol{G}(\boldsymbol{X})^{\intercal}\boldsymbol{P}\boldsymbol{G}(\boldsymbol{X})) < \lambda_{max}(\frac{1}{n}\boldsymbol{G}(\boldsymbol{X})^{\intercal}\boldsymbol{P}\boldsymbol{G}(\boldsymbol{X})) \leqslant K_{max}$$

as $n \to \infty$.

This assumption requires non-vanishing explanatory power of the B-splines bases $\boldsymbol{\Phi}(\boldsymbol{X})$ on the factor loading matrix $\boldsymbol{G}(\boldsymbol{X})$.

Assumption 14. *ε_{it} is realized i.i.d. idiosyncratic shocks with $E(\varepsilon_{it}) = 0$ and $\text{var}(\varepsilon_{it}) = \sigma^2$.*

Heteroskedasticity is caused by γ_t, namely, $\text{var}(\gamma_i + \varepsilon_{it}) = \sigma_i^2$.

3.6.2 Main Results

Theorem 4. *Let $\hat{\boldsymbol{F}}$ be the $T \times J$ matrix estimate of latent risk factors. Under Assumption 10-13, $\hat{\boldsymbol{F}} \to^P \boldsymbol{F}$, as $n \to \infty$.*

Theorem 5. *Define the $n \times J$ matrix $\hat{\boldsymbol{G}}(\boldsymbol{X})$ as the estimate of factor loadings $\boldsymbol{G}(\boldsymbol{X})$. Under Assumption 10-13 and Theorem 5 , as $n \to \infty$, then $\hat{\boldsymbol{G}}(\boldsymbol{X}) \to^P \boldsymbol{G}(\boldsymbol{X})$.*

Theorem 6. *Let the $PH_n \times 1$ vector $\hat{\boldsymbol{A}}$ be the solution of constrained OLS Equation 3.6, then*

$$\hat{\boldsymbol{A}} = \boldsymbol{Q}\tilde{\boldsymbol{A}},$$

where

$$\boldsymbol{Q} = \boldsymbol{I} - (\boldsymbol{\Phi}(\boldsymbol{X})^{\intercal}\boldsymbol{\Phi}(\boldsymbol{X}))^{-1}\boldsymbol{\Phi}(\boldsymbol{X})^{\intercal}\hat{\boldsymbol{G}}(\boldsymbol{X})(\hat{\boldsymbol{G}}(\boldsymbol{X})^{\intercal}\hat{\boldsymbol{G}}(\boldsymbol{X}))^{-1}\hat{\boldsymbol{G}}(\boldsymbol{X})^{\intercal}\boldsymbol{\Phi}(\boldsymbol{X}),$$

$$\tilde{\boldsymbol{A}} = \frac{1}{T}(\boldsymbol{\Phi}(\boldsymbol{X})^{\intercal}\boldsymbol{\Phi}(\boldsymbol{X}))^{-1}\boldsymbol{\Phi}(\boldsymbol{X})^{\intercal}(\boldsymbol{Y} - \hat{\boldsymbol{G}}(\boldsymbol{X})\hat{\boldsymbol{F}}^{\intercal})\boldsymbol{1}_T^{\intercal}.$$

Under Assumption 10-13, $\boldsymbol{\Phi}(\boldsymbol{X})\hat{\boldsymbol{A}} \to^P h(\boldsymbol{X})$, as $n \to \infty$.

Theorem 7. *Under Assumption 12 and Assumption 14, $E(Z|\boldsymbol{H}_0) = \sqrt{2\log PH_n}$.*

Theorem 8. *Under $n \to \infty$ and H_0, given the properties of S_0 and S_1, then:*

$$S \to^d N(0,1),$$

the power of S is approaching 1 *once the arbitrage characteristic is selected as:*

$$\Pr(\textit{reject } H_0 | \hat{\mathscr{M}} \neq \emptyset) \to 1.$$

3.7 Numerical Study

In this section, we use Compustat and Fama-French three and five factors data to simulate stock returns and then evaluate the performance of our estimation and hypothesis test procedures.

3.7.1 Data Generation

Firstly, we use Fama-French three factors monthly returns and all the characteristics that will be included in the empirical study to simulate the stock excess returns. Most of the characteristics are updated annually so we treat those variables as time-invariant during each one-year rolling block. For the characteristics that are updated every month, we substitute the mean values as their fixed values for each fiscal year. We use Fama-French monthly returns from July of year t to June of year $t+1$ and characteristics of fiscal year $t-1$ to generate the stock returns from July of year t to June of year $t+1$. The periods we generate are the same as the empirical study, namely, 50 years from July 1967 to June 2017. For each rolling block with 12 months we have:

$$y_{it} = h(X_i) + \sum_{j=1}^{3} g_j(\boldsymbol{X}_{ij}) f_{jt} + \varepsilon_{it}, \tag{3.7}$$

y_{it} is the generated stock's return; $h(X_i)$ is the mispricing function consisting of a nonlinear characteristic function of x_i, which is to mimic the sparse structure of the mispricing function; $g_j(\boldsymbol{X}_{ij})$ is the j^{th} characteristics-based factor loading, which has an additive semiparametric structure; $\boldsymbol{X}_{ij}$ is the j^{th} subset consisting of 4 characteristics; f_{jt} is the j^{th} Fama-French factor returns at time t; ε_{it} is the idiosyncratic shock for stock i at time t, generated from $N(0,\sigma^2)$.

We generate characteristic functions:

$$h(X_i) = \sin X_i,$$

$$g_j(\boldsymbol{X}_{ij}) = X_{ij1}^2 + (3X_{ij2}^3 - 2X_{ij2}^2) + (3X_{ij3}^3 - 2X_{ij3}) + X_{ij4}^2,$$

X_{ijl} is a randomly picked characteristic without replacement from the data in empirical study and $j = 1,2,3$, $l = 1,\ldots,4$. A description of these characteristics can be found in the Appendix. Additionally, all $h(X_i)$, $g_j(\boldsymbol{X}_{ij})$ are rescaled to have zero mean and unit variance. As we use real data to conduct the simulation, the Assumption of independent X_i may not be satisfied. Although some characteristics are correlated, the semiparametric model overcomes this problem properly when compared with the parametric model that has serious size distortion.

We do not specify $h(X_i)$ and $g_j(\boldsymbol{X}_{ij})$ to be orthogonal explicitly, but we draw characteristics without replacement and employ sine-waves and polynomials to approximate the orthogonality as much as possible. In this simulation, our method can only estimate the component of $h(X_i)$ that is orthogonal to $g_j(\boldsymbol{X}_{ij})$. However, results reveal that one can still select the arbitrage characteristics even if we cannot estimate arbitrary $h(X_i)$ unrestrictively.

3.7.2 Model Misspecification

In this subsection, we show the necessity to consider semiparametric analysis when the forms of factor loadings and mispricing functions are nonlinear.

Under the data generation process, we consider both semiparametric and linear analysis to compare Mean Squared Error (MSE) and hypothesis test results under both specifications. We apply our estimation methodology in section 3.3 to estimate Equation 3.7. For semiparametric specification, we choose the number of B-splines bases to be $\lfloor n^{0.3} \rceil$. n is the number of assets in each balanced rolling window, and $\lfloor \cdot \rceil$ means the nearest integer. We orthogonalize these bases, and then use the Projected-PCA and restricted OLS to estimate model Equation 3.7. As for the hypothesis test part, we choose threshold value to be $\eta_n = H_n\sqrt{2\log(PH_n)} = \lfloor n^{0.3} \rceil \sqrt{2\log(P\lfloor n^{0.3} \rceil)}$, where P is the number of characteristics, and n is the number of stock in each rolling block. For the linear specification, each characteristic only has one basis, which is itself. In terms of the hypothesis test, we use the same logic as in the semiparametric settings, and we set $\eta_n = \sqrt{3\log(P)}$.

In all the estimation above, we assume that we know the real number of factors, which is three. We will discuss the situation when the number of factors is unknown in the next subsection. Mean Squared Error (MSE) is also reported to compare the fitness of models Equation 3.7.

From Table 3.1, under different noise levels, namely $\sigma^2 = 1$ and $\sigma^2 = 4$, the semiparametric model outperforms the linear model in the following aspects:

1 The fitness of the semiparametric model is much better than the linear model, which can be illustrated from MSE.

2 The semiparametric model can enhance the power of S_1 by non-zero S_0, which can not only select the correct mispricing characteristics but also avoid size distortions. As for the linear model, it is influenced by the correlated characteristics. Therefore, during certain periods we even obtain the non-invertible characteristic matrix. The linear model can also select the relevant covariates with decent probability, but it suffers from serious size distortions. In contrast, our semiparametric model with orthogonal bases can mitigate this problem to a great extent.

3 Because S_1 can be very small and even negative, especially when the noise σ_i is strong, the additional component S_0 is necessary to strengthen the power of S_1 and to select the relevant characteristics that can explain the mispricing function.

Table 3.1 Simulation Results 1 Part1

		$\sigma^2 = 1$												$\sigma^2 = 4$									
		Linear Model						semiparametric Model						Linear Model						semiparametric Model					
Window	n	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%
1	468	24.9	11.5	13.4	6.4	100%	100%	-0.5	6.2	-5.7	6	81.2%	0%	14.2	10.8	3.4	8.6	100%	87.4%	-8.2	0	-8.2	8.1	0%	0%
2	894	32.8	11.6	21.2	2	100%	100%	3.4	8	-4.6	1.6	99.9%	0%	11.4	5.8	5.6	4.3	100%	2.1%	-8.5	0	-8.5	3.7	0%	0%
3	1108	34.4	5.7	28.7	11.9	100%	0%	8.6	9	-0.4	11.5	100%	0%	17.1	5.7	11.4	14.1	100%	0%	-7	0.7	-7.7	13.7	7.3%	0%
4	1199	-0.57	0	-0.57	10.2	0%	0%	9.2	9.1	0.1	9.5	96.8%	4.3%	-1.4	0	-1.4	12.5	0%	0%	-6.1		0.06	-6.2	7%	0%
5	1333	92	19.6	72.4	2.31	100%	100%	10.6	9	1.6	2	100%	0%	28.2	6.1	22	4.5	100%	6.5%	0.2	7.4	-7.2	4.1	82.8%	0%
6	1409	90	28.5	61.5	16	100%	100%	28.6	12.6	15.9	15.8	100%	28%	45.3	16.1	29.2	18.4	100%	73.4%	16.3	10.9	5.4	17.5	68.4%	35.9%
7	1466	78.4	10.6	67.8	6.4	100%	74.2%	19.5	9	10.5	6.2	100%	0%	34.8	5.7	29.1	8.6	100%	0.02%	4.3	9	-4.7	8.4	99.9%	0%
8	1560	133	16.8	116.2	3.3	100%	100%	20.3	10	10.3	3.2	100%	0%	45.2	6.1	39.1	5.5	100%	6.9%	4.2	10	-5.8	5.4	100%	0%
9	1494	117.7	13.6	104.1	3.6	100%	100%	23.1	9	14.1	3.5	100%	0%	44.1	7.6	36.5	5.8	100%	32.4%	6	9	-3	5.6	100%	0.1%
10	1292	90.7	11.5	79.2	3.7	100%	100%	16.2	9	7.2	3.6	100%	0%	39.5	9.3	30.2	5.9	100%	61.1%	3.6	8.9	-5.3	5.7	99.7%	0%
11	1393	84.7	10.6	74.1	6.1	100%	85.1%	20.7	9.1	11.6	5.8	100%	1.1%	37.1	6.5	30.6	8.3	100%	12.9%	8.9	8.9	0	7.8	98.1%	1.3%
12	1340	83.5	28	55.5	2.38	100%	100%	10.6	9	1.6	2	100%	0%	26	6.2	19.8	4.6	100%	7.1%	-1.8	5.7	-7.5	4.1	63.7	0%
13	1285	113.8	16	97.8	1.73	100%	100%	10.6	9	1.6	1.6	100%	0%	34.5	6.6	27.9	4	100%	15.3%	-2.4	5.1	-7.5	3.7	57.1%	0%
14	1181	88.5	12.8	75.7	4.7	100%	100%	15.8	9	6.8	4.5	100%	0%	31.2	5.9	25.3	6.9	100%	2.3%	3.7	9	-5.3	6.6	100%	0%
15	1110	45.7	7.5	38.1	8.9	100%	30.4%	11.5	9	2.5	8.7	100%	0%	23.9	5.8	18.1	11.1	100%	0.6%	-2	4.8	-6.8	10.8	0.54%	0%
16	1044	20.5	5.7	14.8	18.4	100%	0%	9.9	9	0.9	17.9	100%	0%	14.6	5.7	8.9	20.6	100%	0%	1.2	6.1	-4.9	20	68.1%	0.2%
17	1125	59.4	11.5	47.9	9.2	100%	100%	13.2	9	4.2	9	100%	0%	27.2	6.2	21	11.5	100%	8.4%	2.6	8.8	-6.2	11	97.9%	0%
18	2192	NA	NA	NA	NA	NA	NA	23.2	11	12.2	4.3	100%	0%	NA	NA	NA	NA	NA	NA	6.7	11	-4.3	6.4	100%	0%
19	2236	56.1	11.5	44.6	5.8	100%	100%	17.8	11	6.8	5.2	100%	0%	28.3	6.3	22	8	100%	20.3%	4.3	11	-6.7	7.4	100%	0%
20	2273	43.3	5.7	37.6	3.8	100%	0%	22.4	11	11.4	3.2	100%	0%	22.4	5.7	16.7	6.1	100%	0%	5	10.2	-5.2	5.4	92.6%	0%
21	2235	59.8	11.8	48	2.7	100%	100%	20.2	11	9.2	2	100%	0%	25	7.3	17.7	4.9	100%	28.2%	4.6	11	-6.4	4.2	100%	0%
22	2270	40.2	11.5	28.7	2.78	100%	99.5%	17.2	11.6	5.6	2.1	100%	0%	17.1	5.9	11.2	5	100%	3.5%	-6	0.1	-6.1	4.2	1.1%	0%
23	2405	41.4	8.9	32.5	4.1	100%	54.2%	16.3	11	5.3	3.3	100%	0%	18.7	5.8	12.9	6.3	100%	7.1%	-3.3	3	-6.3	5.5	27.3%	0%
24	2376	19	9.7	9.3	1.8	100%	69.9%	23.1	11	12.1	1	100%	0%	7.5	5.7	1.8	4	100%	0%	5.6	11	-5.4	3.2	100%	0%
25	2323	15.9	9.5	6.4	3.5	66.7%	98.6%	20.6	11	9.6	2.7	100%	0%	1.1	0	1.1	5.8	0%	0%	5.3	11	-5.7	4.9	100%	0%
26	2344	NA	NA	NA	NA	NA	NA	24.9	12.9	12	3.3	100%	17.1%	NA	NA	NA	NA	NA	NA	6.5	11	-4.5	5.4	100%	0%
27	2434	NA	NA	NA	NA	NA	NA	27.3	11	16.3	1.2	100%	0%	NA	NA	NA	NA	NA	NA	6.9	11	-4.1	3.4	100%	0%
28	2548	0.9	0	0.9	4.2	0%	0%	26.2	11	15.2	3.3	100%	0%	-1.3	0	-1.3	6.5	0%	0%	6.9	11	-4.1	5.5	100%	0%
29	2741	10.3	5.7	4.5	4.2	100%	0%	58.2	11.1	47.1	3.4	100%	1.3%	6.6	5.7	0.9	6.4	100%	0%	17.6	11	6.6	5.5	100%	0%
30	2928	5.6	4.6	1	7.1	80.4%	0%	59.2	11.8	47.4	6.3	100%	7.8%	-0.4	0.1	-0.5	9.3	2.5%	0%	18.8	11	7.8	8.5	100%	0.3%
31	2894	13.4	5.7	7.7	6.4	100%	0%	61	13.4	47.6	5.7	100%	21.6%	8.1	5.7	2.3	8.6	100%	0%	17.7	11	6.7	7.8	100%	0.2%
32	2905	23.1	11.5	11.6	5.9	100%	100%	33.2	11.3	21.9	5.2	100%	3%	12.9	8.5	4.4	8.1	100%	48.2%	9.8	11	-1.2	7.4	100%	0%
33	2804	9.8	5.7	4.1	9.6	100%	0%	42.7	18.5	24.2	8.9	100%	68.5%	7.3	5.7	1.6	11.9	100%	0%	9.7	11	-1.3	11.2	100%	0%
34	2570	6.9	5.7	1.2	22	99.7%	0%	37.3	12.2	25.1	21.2	100%	10.4%	2	1.9	0.1	24	34.4%	0%	12.7	11	1.7	23.3	100%	0.2%
35	2516	8.3	5.7	2.6	7.9	100%	0%	41.3	11	30.3	7.2	100%	0.4%	5.1	5.02	0.08	10.1	87.3%	0%	12.9	11	1.9	9.4	100%	0%
36	2491	10.7	5.7	4.9	2.1	100%	0%	41.3	11	30.3	1.4	100%	0.4%	0.5	0.25	0.25	4.4	4.5%	0%	12.4	11	1.4	3.6	100%	0%
37	2402	14.1	5.7	8.4	5.6	100%	0%	26.5	11.2	15.3	4.9	100%	2.2%	8.8	5.7	3.1	7.9	100%	0%	7.9	11	-3.1	7.1	100%	0%
38	2326	19.7	9.6	10.1	3	100%	66.8%	28.9	11.3	17.6	2.3	100%	2.1%	8.1	5.8	2.3	5.3	100%	0.3%	8.7	11	-2.3	4.4	99.9%	0.1%
39	2241	17	5.7	16.1	2.9	100%	0.2%	11	11	0	1.7	100%	0%	9.1	5.8	2.3	5.3	100%	0.3%	-7.5	0.1	-7.6	4	1.1%	0%
40	2178	21.8	5.7	16.1	2.9	100%	0%	9.5	11	-1.5	2.2	100%	0.3%	12.2	5.7	6.5	5.2	100%	0%	-8.1	0	-8.1	4.4	0%	0%

Table 3.2 Simulation Results 1 Part2

		$\sigma^2 = 1$												$\sigma^2 = 4$									
		Linear Model						semiparametric Model						Linear Model						semiparametric Model					
Window	n	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%
41	2113	24.1	6.1	18	4.7	100%	7.5%	7.8	10	-2.2	3.9	100%	0%	13.9	5.7	8.2	6.9	100%	0%	-8.1	0	-8.1	6.1	0%	0%
42	2023	18.4	5.7	12.7	6.8	100%	0%	11.3	10	1.3	6	100%	0%	10.8	5.8	5.1	9	100%	0%	-7.1	0.3	7.4	8.2	2.7%	0%
43	2007	18.8	5.7	13.1	4.9	100%	0%	9.1	10	-0.9	4.1	100%	0%	10.5	5.7	4.8	7.1	100%	0%	-8.3	0	-8.3	6.3	0%	0%
44	1924	16.6	5.8	10.8	8.18	100%	0.2%	13.6	10.8	2.8	7.5	100%	8%	11.2	5.8	5.4	10.4	100%	0.3%	-3.5	2.7	-6.2	9.7	26.3%	0.2%
45	1990	27.5	5.7	21.8	2.1	100%	0%	8.1	10	-1.9	1.4	100%	0%	13.3	5.7	7.5	4.4	100%	0%	-8	0	-8	3.6	0%	0%
46	1937	20.3	5.8	14.5	5.4	100%	0.9%	19.7	11.8	7.9	4.7	100%	18%	12.6	5.9	6.7	7.6	100%	3%	8	11.2	-3.2	6.8	100%	12.3%
47	1909	13.2	5.7	7.5	5.2	100%	0%	14.2	10.4	3.8	4.5	100%	3.5%	8.8	5.7	3.1	7.4	100%	0%	2.7	8.4	-5.7	6.7	84.9%	0%
48	1872	21.8	5.7	16.1	2.7	100%	0%	11.4	10	1.4	2	100%	0%	11.1	5.8	5.3	4.9	100%	0%	-6.8	0.6	-7.4	4.2	5.7%	0%
49	1841	16.3	5.7	10.5	2.1	100%	0%	8.7	10	-1.3	1.4	100%	0.1%	8.1	5.7	2.4	4.4	100%	0%	-8.4	0	-8.4	3.6	0%	0%
50	1826	11	5.7	5.3	4.3	100%	0%	12.6	10.6	2	3.5	100%	3.5%	6.5	5.7	0.8	6.6	99.7%	0.3%	-6.9	0	-6.9	5.7	0%	0%

This table documents results under the characteristics-based beta and alpha of Fama-French 3 factors model. To mimic the empirical study, we simulated 50 12-month rolling windows, and each window is repeated for 1000 times. Each column summarises the mean value of 1000 estimations and test results. S_1 is the conventional Wald test while S_0 is the power-strengthened component. This table also compares the performance of both semiparametric and linear models under different noise levels, $\sigma^2 = 1$ and $\sigma^2 = 4$. NA results are caused by non-invertible characteristic matrices. "Selected" means the percentage of selecting the relevant characteristic in the mispricing function in 1000 experiments. Similarly, "distortion" represents the percentage of wrongly selecting irrelevant characteristics in 1000 repetitions.

3.7.3 Robustness Under Stronger Noise

In Table 3.1, we set two different noise levels of random shocks, namely $\sigma^2 = 1$ and $\sigma^2 = 4$. Although $\sigma^2 = 1$ is closer to the empirical data, we conduct this comparison to show the robustness of our methods. When the noise level becomes three times bigger, the accuracy of power-enhanced tests gets much lower for certain windows. However, there are no size distortions under comparatively high noise level recalling that all the components of our simulation model are rescaled to have unit variance. Another fact is that the stronger noise does deteriorate the power of conventional Wald tests, leading to an even smaller value of S_1, which can be mitigated through adding S_0.

Therefore, we conclude that our methods are robust to a higher noise level regarding no size distortions. However, the accuracy of selecting relevant components and the role of enhancing the power of hypothesis tests will be influenced negatively.

3.7.4 Number of Factors

In the empirical study, the number of factors is unknown. Therefore, in this subsection we will study whether our methodology is robust to a various numbers of factors considered.

We simulate according to another data generation process:

$$y_{it} = h(X_i) + \sum_{j=1}^{5} g_j(\boldsymbol{X}_{ij}) f_{jt} + \varepsilon_{it}, \tag{3.8}$$

similarly, y_{it} is the generated stock return; $h(X_i)$ is the mispricing function consisting of a nonlinear characteristic function of X_i, to mimic the sparse structure of the mispricing function; $g_j(\boldsymbol{X}_{ij})$ is the j^{th} characteristics-based factor loading, which has an additive semiparametric structure; X_{ij} is a subset consisting of four characteristics; f_{jt} is the j Fama-French 5-factor returns at time t; ε_{it} is the idiosyncratic shock, generated from $N(0, \sigma^2)$. Moreover, we generate characteristic functions as:

$$h(X_i) = \sin X_i,$$

$$g_j(\boldsymbol{X}_j) = X_{ij1}^2 + (3X_{ij2}^3 - 2X_{ij2}^2) + (3X_{ij3}^3 - 2X_{ij3}) + X_{ij4}^2,$$

where X_{ijl} is a randomly picked characteristic without replacement from the data in empirical study with $j = 1, \ldots, 5$, $l = 1, \ldots, 4$. Furthermore, all $h(X_i)$ and $g_j(\boldsymbol{X}_{ij})$ are rescaled to have zero mean and unit variance.

Given the above data generation process, together with the data generation process, we test the influence of over and under-estimated number of factors. We choose the number of factors to be either three or five, and compare the results in Table 3.3. The first category column is the scenario of over estimating the number of factors. We simulate the data generation process using the Fama-French three factors model, but estimate the number of factors to be five. However, this does not cause any serious problems. For some rolling blocks, the probability of mistakenly selected irrelevant characteristics is slightly higher under over estimating the number of factors. Moreover, over estimating the number of factors can increase the model fitting marginally. Therefore, we conclude that over estimating the number of factors does not cause severe size distortion using our methods.

On the other hand, under estimating the number of factors can lead to misleading test results. We can conclude this from the last column where we estimate the number of factors to be three in a five-factor model. Compared with the correct specified model, under estimating causes not only higher MSE, but also higher distortions, which means it is more likely to select irrelevant characteristics. Therefore, in the empirical study we prefer the five-factor model rather than the three-factor model.

Table 3.3 Simulation Results 2 Part1

		Number of factors $J=3$												Number of factors $J=5$									
		Number of estimated factors $\hat{J}=5$						Number of estimated factors $\hat{J}=3$						Number of estimated factors $\hat{J}=5$						Number of estimated factors $\hat{J}=3$					
Window	n	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%
1	468	2.6	7	-4.4	5.9	99.6%	0%	-0.5	6.2	-5.7	6	81.2%	0%	4.5	6.9	-2.4	6	97.4%	1.3%	-8.5	0	-8.5	6.9	0%	0%
2	894	6	8	-2	1.5	99.9%	0%	3.4	8	-4.6	1.6	99.9%	0%	5.5	8	-2.5	2.3	100%	0%	-6.5	1	-7.5	3	12.9%	0%
3	1108	12.8	9	3.8	11.4	100%	0.1%	8.6	9	-0.4	11.5	100%	0%	14.3	9	5.3	13.6	100%	0.5%	5.3	9	-3.7	14.1	100%	0.1%
4	1199	13.5	10.3	3.2	9.4	99.8%	0%	9.2	9.1	0.1	9.5	96.8%	4.3%	16.6	11.7	4.9	9.8	99.5%	25.6%	2.3	3	-0.7	10.1	23.5%	10%
5	1333	15.4	9	6.4	1.8	100%	0.1%	10.6	9	1.6	2	100%	0%	16.7	9	7.6	2.7	100%	0%	4.5	9	-4.5	3.4	100 %	0%
6	1409	41.6	17.5	24.1	15.6	100%	51.3%	28.6	12.6	15.9	15.8	100%	28%	58.7	28.3	30.4	13.5	100%	90%	106.1	29.9	76.2	13.2	100%	100%
7	1466	26.8	9	17.8	6.1	100%	0.01%	19.5	9	10.5	6.2	100%	0%	26.3	9	17.3	9.2	100%	0.3%	3.5	9	-5.5	11.7	100%	0%
8	1560	27.6	10	17.6	3	100%	0%	20.3	10	10.3	3.2	100%	0%	30.4	10	20.4	5	100%	0.5%	26.7	24	2.7	6.7	100%	100%
9	1494	31.7	9.1	22.6	3.3	100%	0.7%	23.1	9	14.1	3.5	100%	0%	32.1	9.2	22.9	4.4	100%	1.4%	29.1	18	11.1	4.6	100%	100%
10	1292	22.5	9	13.5	3.4	100%	0.1%	16.2	9	7.2	3.6	100%	0%	26.3	10	16.3	4.3	100%	11.3%	46.7	18	28.7	4.4	100%	100%
11	1393	27.8	9.4	18.4	5.7	100%	4%	20.7	9.1	11.6	5.8	100%	1.1%	30	10.7	19.3	5.6	100%	17.2%	49	29.1	19.9	5.8	100%	100%
12	1340	15.2	9	6.2	1.8	100%	0%	10.6	9	1.6	2	100%	0%	15.2	9	6.2	1.8	100%	0%	4	9	-5	2.7	100%	0%
13	1285	15.4	9	6.4	1.4	100%	0.2%	10.6	9	1.6	1.6	100%	0%	15.1	9	6.1	1.4	100%	0.1%	3.5	9	-4.5	2.5	100%	0%
14	1181	21.9	9	12.9	4.4	100%	0.2%	15.8	9	6.8	4.5	100%	0%	21.4	9	12.4	4.7	100%	0.2%	4.5	9	-4.5	6	100%	0%
15	1110	16.4	9	7.4	8.5	100%	0%	11.5	9	2.5	8.7	100%	0%	17.1	9	8.1	9.8	100%	0.1%	5.3	9	-3.7	10.2	100%	0%
16	1044	13.3	9.1	4.3	17.8	100%	0.8 %	9.9	9	0.9	17.9	100%	0%	14.9	9.2	5.7	17.8	100%	2.1%	40	22.4	17.6	16.8	100%	100%
17	1125	18.7	9	9.7	8.8	100%	0.1%	13.2	9	4.2	9	100%	0%	24.1	9.7	14.4	10.7	100%	7.1%	101.4	27	74.4	10.3	100%	100%
18	2192	31.8	11	20.8	4.1	100%	0.2%	23.2	11	12.2	4.3	100%	0%	69.8	28.6	41.2	5.4	100%	77.6%	563.8	33	530.8	3.6	100%	100%
19	2236	24.4	11	13.4	5.1	100%	0%	17.8	11	6.8	5.2	100%	0%	25.1	11	14.1	5.4	100%	0%	10.2	11	-0.8	6.1	100%	0%
20	2273	29.4	11	18.4	3	100%	0.4%	22.4	11	11.4	3.2	100%	0%	30.3	11.1	19.2	4.2	100%	0.7%	61.1	33	28.1	5.3	100%	100%
21	2235	27.5	11	16.5	1.8	100%	0%	20.2	11	9.2	2	100%	0%	29	11	18	2.2	100%	0.3%	5.9	11	-5.1	3.3	100%	0%
22	2270	24.9	13.7	11.2	1.9	100%	23.9%	17.2	11.6	5.6	2.1	100%	0%	43.2	20.4	22.8	2.3	100%	56.7%	41.6	22.1	19.5	2.1	100%	100%
23	2405	22.5	11	11.5	3.2	100%	0.1%	16.3	11	5.3	3.3	100%	0%	21.7	11	10.7	3.3	100%	0%	10.9	11.9	-1	4.3	100%	7.8%
24	2376	30.6	11	19.6	0.8	100%	0.1%	23.1	11	12.1	1	100%	0%	30.3	11	19.3	1.2	100%	0%	20.4	21.4	-1	2.7	100%	94.8%
25	2323	27.2	11.1	16.1	2.5	100%	0.4%	20.6	11	9.6	2.7	100%	0%	26.8	11	15.8	2.8	100%	0%	8.5	11	-2.5	3.9	100%	0%
26	2344	36.4	16.7	19.7	3.1	100%	51.3%	24.9	12.9	12	3.3	100%	17.1%	36.1	17	19.1	3.2	100%	54%	47.5	23.3	24.2	4.3	100%	100%
27	2434	36.3	11.1	25.2	1	100%	0.9%	27.3	11	16.3	1.2	100%	0%	38.3	11.3	27	1.3	100%	2.6%	89.5	33	56.5	1.7	100%	100%
28	2548	34.5	11	23.5	3.2	100%	0.1%	26.2	11	15.2	3.3	100%	0%	34.8	11	23.8	3.3	0%	0.2%	50.3	22	28.3	4	100%	100%
29	2741	73	12.3	60.7	3.2	100%	10.9%	58.2	11.1	47.1	3.4	100%	1.3%	79.4	15.4	64	3.5	100%	36.8%	439.7	62.7	377	3.6	100%	100%
30	2928	73.9	13.8	60.1	6.1	24.1%	0%	59.2	11.8	47.4	6.3	100%	7.8%	84.6	18.7	65.9	7.4	100%	52.2%	94	32.6	61.4	7.2	100%	100%
31	2894	77.3	16.3	61	5.5	100%	45.4%	61	13.4	47.6	5.7	100%	21.6%	77.2	16.3	60.9	5.5	100%	45.9%	28.6	11	17.6	6.5	100%	0%
32	2905	42.4	12.9	29.5	5	100%	16%	33.2	11.3	21.9	5.2	100%	3%	41.7	12.8	28.9	6.1	100%	15.7%	8.7	11	-2.3	9.4	100%	0%
33	2804	53.8	20.5	33.3	8.8	100%	86.8%	42.7	18.5	24.2	8.9	100%	68.5%	54.1	20.4	33.6	10.1	100%	85.6%	35.5	22	13.5	12.3	100%	100%
34	2570	47.6	14.2	33.4	21.1	27.2%	0%	37.3	12.2	25.1	21.2	100%	10.4%	49.4	14.5	34.9	41.2	100%	28.9%	53.8	22	31.8	38.8	100%	100%
35	2516	50.9	11.3	39.6	7	100%	2.9%	41.3	11	30.3	7.2	100%	0.4%	38.4	11	27.4	18.4	100%	0.4%	51.2	33	18.2	20.8	100%	100%
36	2491	51.3	11.8	39.5	1.3	100%	6.8%	41.3	11	30.3	1.4	100%	0.4%	50.5	11.3	39.2	1.6	100%	3%	15.9	11	4.9	3.3	100%	0%
37	2402	34.4	12.2	22.2	4.7	100%	10.5%	26.5	11.2	15.3	4.9	100%	2.2%	37.4	14.2	23.2	5.1	100%	29.2%	68.8	22	46.8	6.1	100%	0%
38	2326	37.4	12.3	25.1	2.1	100%	10.3%	28.9	11.3	17.6	2.3	100%	2.1%	37.2	12.2	25	2.8	100%	9.4%	44.6	22	22.6	3.4	100%	0%
39	2241	14.8	11	3.8	1.6	100%	0.1%	11	11	0	1.7	100%	0%	14.9	11	3.9	1.7	100%	0%	23.4	22	1.4	2.4	100%	100%
40	2178	13.1	11.1	2	2	100%	1.1%	9.5	11	-1.5	2.2	100%	0.3%	12.9	11.2	1.8	2.2	100%	1.3%	20.4	13.1	7.3	3.4	13.3%	100%

Table 3.4 Simulation Results 2 Part2

		Number of factors $J=3$								Number of factors $J=5$							
		Number of estimated factors $\hat{J}=5$					Number of estimated factors $\hat{J}=3$					Number of estimated factors $\hat{J}=5$					Number of estimated factors $\hat{J}=3$				
Window	n	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%	S	S_0	S_1	MSE	Selected %	Distortion%
41	2113	11	10	1	3.8	100%	0.2%	7.8	10	-2.2	3.9	100%	0%	11.5	10	1.5	4.5	100%	0.2%	41.1	32.4	8.7	5.4	99.9%	100%
42	2023	15.2	10	5.2	5.9	100%	0%	11.3	10	1.3	6	100%	0%	15.7	10	5.7	6.4	100%	0%	-8	0	-8	9.2	0%	0%
43	2007	12.6	10	2.6	4	100%	0.5%	9.1	10	-0.9	4.1	100%	0%	13.4	10.2	3.2	4.7	100%	1.7%	-0.1	6.4	-6.5	5.6	64.4%	0%
44	1924	19.9	13.1	6.8	7.3	100%	30.7%	13.6	10.8	2.8	7.5	100%	8%	19.5	12.9	6.6	7.5	100%	28.9%	20	20	0	8.3	100%	100%
45	1990	11.4	10	1.4	1.2	100%	0.1%	8.1	10	-1.9	1.4	100%	0%	20.7	14.6	6	1.8	100%	45.2%	116	20	96	1.7	100%	100%
46	1937	27.1	14	13.1	4.5	100%	37.7%	19.7	11.8	7.9	4.7	100%	18%	28.3	14.8	13.5	5.4	100%	45.8%	24.6	20	4.6	6.2	100%	100%
47	1909	19.5	11.7	7.8	4.4	100%	16.1%	14.2	10.4	3.8	4.5	100%	3.5%	24	14	10	4.4	100%	38.1%	51.7	35.2	16.5	5.4	100%	100%
48	1872	15.2	10	5.1	1.8	100%	0.2%	11.4	10	1.4	2	100%	0%	15	10	5	2.1	100%	0.1%	5	10	-5	2.9	100%	0%
49	1841	12.3	10.1	2.2	1.2	100%	1.1%	8.7	10	-1.3	1.4	100%	0.1%	11.8	10.1	1.7	4.4	100%	0.8%	-10	0	-10	4.4	0%	0%
50	1826	18.5	12.4	6.1	3.3	100%	15%	12.6	10.6	2	3.5	100%	3.5%	20.2	13.3	6.9	3.7	100%	19.3%	-3.9	0.4	-4.3	4.4	3.9%	0%

This table presents results under the characteristics-based beta and alpha of both Fama-French 3 and 5 factors models. To mimic the empirical study, we simulated 50 12-month rolling windows, and each window is repeated for 1000 times. Each column summarises the mean value of 1000 estimation and test results. We compare the results of both over and under estimating the number of factors, namely,$\hat{J}=3$ and $\hat{J}=5$. S_1 is the conventional Wald test while S_0 is the power-strengthened component. NA results are caused by non-invertible characteristic matrices. "Selected" means the percentage of selecting the relevant characteristic in mispricing functions in 1000 repetitions. Similarly, "distortion" represents the percentage of wrongly selecting irrelevant characteristics in 1000 experiments.

3.8 Empirical Study

3.8.1 Data

We use monthly stock returns from CRSP and firms' characteristics from Compustat, ranged from 1965 to 2017. We construct 33 characteristics following the methods of Freyberger et al. (2020). Details of these characteristics can be found in the Appendix. We use characteristics from fiscal year $t-1$ to explain stock returns between July of year t to June of year $t+1$. After adjusting the dates from the balance sheet data, we merge two data sets from CRSP and Compustat. We require all firms included in our analysis to have at least three years of characteristics data in Compustat.

Data is modified with regards to the following aspects:

1 Delisting is quite common for CRSP data. We use the way of Hou et al. (2015) to correct the returns of all delisted assets before 2018. Detailed methods can be found in their Appendix.

2 Projected-PCA works well, even under small T circumstances. Thus, we choose the width of our window to be 12 months. Another reason for the short window width is that we assume that mispricing functions are time-invariant in each window. One of the limitations of Projected-PCA is that it can only be used for a balanced panel, which means the number of stock will vary when we applied one-year rolling windows to obtain a short-time balanced panel. Meanwhile, we take monthly updated characteristics' mean values of 12 months as fixed characteristic values in each window. We also use the rolling window method to detect peer groups of arbitrage characteristics.

3 B-splines are based on each time-invariant characteristic of n firms, which are not delisted in each window.

4 Rolling windows are moving at a 12-month step from Jul. 1967 to Jun. 2017 without overlapping. The first 24 months returns are not included as they do not have corresponding characteristics.

5 Excess returns are obtained by the difference between monthly stock returns and monthly Fama-French risk-free returns.

3.8.2 Estimation

We construct B-splines bases based on evenly distributed knots, and the degree of each basis is three. We choose $H_n = \lfloor n^{0.3} \rceil$, and n is the number of stocks. To get a relatively large balanced panel in each window, some characteristics with too many missing values are eliminated. Therefore, only 33 characteristics are left. Firms kept in balanced panels in our dataset range from 468 to 2928, which means that both n and $\hat{\boldsymbol{A}} \in \mathbb{R}^{PH_n}$ are diverging. Large n can satisfy asymptotic requirements. These facts emphasize the necessity of introducing a power-enhanced component into the hypothesis test. Before the next step, we use time-demeaning matrix $\boldsymbol{D}_T$ to demean excess return matrix in each window.

Next, we project the time-demeaned monthly excess return matrix $\tilde{\boldsymbol{Y}}$ onto the B-splines space spanned by characteristics bases $\boldsymbol{\Phi}(\boldsymbol{X})$, and then we collect the fitted values $\hat{\boldsymbol{Y}}$. We apply Principal Component Analysis on $\frac{1}{n}\hat{\boldsymbol{Y}}^\top\hat{\boldsymbol{Y}}$, and attain the first five eigenvectors corresponding to the first five biggest eigenvalues as the estimates of unobservable factors $\boldsymbol{F}$. We choose the number of factors to be five according to simulation results.

Then, we estimate the factor loading matrix by:

$$\hat{\boldsymbol{G}}(\boldsymbol{X}) = \hat{\boldsymbol{Y}}\hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}^\top\hat{\boldsymbol{F}})^{-1}.$$

Moreover, we use equality-constrained OLS to estimate the mispricing function. We project excess monthly return matrix on the characteristic space $\boldsymbol{\Phi}(\boldsymbol{X})$ that is orthogonal to factor loading matrix $\hat{\boldsymbol{G}}(\boldsymbol{X})$.

Another goal of this paper is to conduct a power-enhanced test on the mispricing function. Therefore, our final step is to estimate the covariance matrix $\boldsymbol{\Sigma}$ of $\hat{\boldsymbol{A}}$.

3.8.3 Power-enhanced Hypothesis Tests

In this section, we conduct a power-enhanced test in each rolling block. Firstly, we set threshold value for each window, $\eta_n = H_n\sqrt{2\log(PH_n)}$, where H_n is the number of bases for each characteristic whereas P is the number of total characteristics in each window, with $P = 33$. η_n is data-driven critical value and it diverges as the number of firms increases. We use indicator function $\boldsymbol{I}(\sum_{h=1}^{H_n}|\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \geqslant \eta_n)$ with critical value $\eta_n = H_n\sqrt{2\log(PH_n)}$ to achieve three goals.

1 This indicator function select the most relevant characteristics that can explain the variation of the mispricing function. Results of last column in Table 3.5 are characteristics selected in $\hat{\mathscr{M}} = \{\boldsymbol{X}_p \in \boldsymbol{X} : \sum_{h=1}^{H_n} |\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \geqslant \eta_n, \quad p = 1,2,\ldots,P\}$.

2 It contributes to the test statistics S by adding a diverging power-enhanced component S_0. As $T = 12$ is small in this empirical study, we assume the homoskedasticity of $\varepsilon_{it} + \gamma_i$. We also specify an over-shrunk covariance matrix by setting off-diagonal elements to be zeros.

3 It avoids size-distortion by the conservative critical value η_n.

The diagonal elements of $\hat{\boldsymbol{\Sigma}}$ are estimated variances of mispricing coefficients. These elements can be substituted into the indicator function $\boldsymbol{I}(\sum_{h=1}^{H_n} |\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \geqslant \eta_n)$, where $\hat{\sigma}_{ph}$ is the ph^{th} diagonal element of $\hat{\boldsymbol{\Sigma}}$.

Finally, the new statistics S can be calculated as:

$$S = S_0 + S_1,$$

$$S_0 = H_n \sum_{p=1}^{P} \boldsymbol{I}(\sum_{h=1}^{H_n} |\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \geqslant \eta_n), \quad S_1 = \frac{\hat{\boldsymbol{A}}^{\intercal}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{A}} - PH_n}{\sqrt{2PH_n}}.$$

3.8.4 Test Results

This section presents the empirical results. Details can be found in Table 3.5, which lists the results of 50 rolling windows from Jul.1967 to Jun.2017. Generally, the number of firms included in the 12-month rolling block is increasing period by period. The number of our characteristic B-splines bases is a function of the number of firms n in each block. Therefore, the dimension of the mispricing coefficient vector $\boldsymbol{A} \in \mathscr{R}^{PH_n}$ is also diverging. This verifies the necessity of using a power-enhanced component S_0.

Recalling that $S|H_0 \to^d N(0,1)$, some of the test statistics S is big enough to reject the null hypothesis. However, for some testing windows, there are no strong signals showing the existence of characteristics-based mispricing functions after subtracting systematic effects. Most S_1 values are small and even negative, which may be caused by the sparsity structure of the mispricing function or/and the low power problems due to diverging dimension of mispricing coefficients.

The power-enhanced component S_0 works well in the empirical study. It selects the most important explanatory characteristics and strengthens the power of S_1, mitigating the low power problem.

Apart from contributing to the power of tests, the indicator function in the power-enhanced component can also screen out the most relevant characteristics, which are concluded as "Characteristics Selected" in Table 3.5.

Some empirical findings are worth discussing. Although short-term cumulative returns like r_{2_1} are always selected, we cannot take this as the evidence of arbitrage opportunities since we construct r_{2_1} as the time-invariant average of all r_{2_1} in the same rolling window, which contains much overlapping information of monthly excess returns. However, this is not the case for long-term and mid-term cumulative returns like r_{12_2}, r_{12_7} and r_{6_2}, because these average returns include a lot of information from another rolling window.

Apart from the cumulative returns, some other characteristics contribute to the arbitrage opportunities as well. PCM (Price to Cost Margin) appears twice. From Figure 2, we find that the PCM mispricing curve is nonlinear and generally decreasing as the value of PCM increases. ROA (Return-On-Asset) also plays a role during 1988-1989. It behaves like a parabola with fluctuations near zero in Figure 3. As for Lev (ratio of long-term debt and debt in the current liabilities), it is decreasing for Lev 0 and increasing afterward as in Figure 7. In Figure 8, IPM (pre-tax Profit Margin) function behaves like a "V" shape with the turning point zero during 2004-2005. DelGmSale (difference in the percentage in Gross margin and the percentage change in Sales) experiences a bump at zero during 2015-2016 in Figure 9. C2D (cash flow to price) curve behaves like "V" around the zero in 2016-2017, (see Figure 10). All characteristics curves in the above figures are standardized as uniformly distributed characteristics in the interval $[-100, 100]$. This is for presentation purposes only since most characteristics are unevenly distributed.

Another finding is the persistence of some arbitrage characteristics. Arbitrage characteristics can be persistent for two years once appeared, such as BEME (ratio of the book value of equity and market value of equity) in Figure 4. Some persistent arbitrage characteristics even have similar shapes of mispricing functions in different rolling windows, such as AT (Total asset) in Figure 6 and LME (total market capitalization of the previous month) in Figure 5.

Table 3.5 Empirical Study Results

Time period	n	S	S_0	S_1	MSE	Characteristics Selected
Jul.1967-Jun.1968	468	-9.6	0	-9.6	0.005	NONE
Jul.1968-Jun.1969	951	-0.45	8	-8.45	0.004	r_{2_1}
Jul.1969-Jun.1970	1108	1.7	9	-7.3	0.005	r_{2_1}
Jul.1970-Jun.1971	1199	-8.7	0	-8.7	0.006	NONE
Jul.1971-Jun.1972	1333	-10	0	-10	0.004	NONE
Jul.1972-Jun.1973	1409	12.7	18	-5.3	0.005	r_{12_2},r_{6_2}
Jul.1973-Jun.1974	1466	2.1	9	-6.9	0.005	r_{2_1}
Jul.1974-Jun.1975	1560	-10.7	0	-10.7	0.01	NONE
Jul.1975-Jun.1976	1494	0.1	9	8.9	0.05	r_{2_1}
Jul.1976-Jun.1977	1292	0.1	9	-9	0.004	r_{2_1}
Jul.1977-Jun.1978	1393	-9.4	0	-9.4	0.005	NONE
Jul.1978-Jun.1979	1340	8.6	18	-9.4	0.005	r_{2_1},r_{12_7}
Jul.1979-Jun.1980	1285	1	9	-8	0.005	r_{2_1}
Jul.1980-Jun.1981	1181	9.7	18	-8.2	0.006	r_{12_7},r_{12_2}
Jul.1981-Jun.1982	1110	1.2	9	-7.8	0.01	r_{2_1}
Jul.1982-Jun.1983	1044	33.1	36	-3	0.01	r_{12_2},r_{12_7},r_{6_2},r_{2_1}
Jul.1983-Jun.1984	1125	-0.9	9	-9.9	0.006	r_{2_1}
Jul.1984-Jun.1985	2192	-0.2	11	-11.2	0.01	r_{2_1}
Jul.1985-Jun.1986	2236	13.1	22	-8.94	0.01	r_{12_7},r_{12_2}
Jul.1986-Jun.1987	2273	1.7	11	-9.3	0.01	PCM
Jul.1987-Jun.1988	2235	0.9	11	-10.1	0.01	r_{2_1}
Jul.1988-Jun.1989	2270	1.2	11	-9.8	0.01	ROA
Jul.1989-Jun.1990	2405	-0.1	11	-11.1	0.01	r_{2_1}
Jul.1990-Jun.1991	2376	1.1	11	-9.9	0.02	r_{2_1}
Jul.1991-Jun.1992	2323	2.1	11	-8.9	0.02	r_{2_1}
Jul.1992-Jun.1993	2344	12.2	22	-9.8	0.02	r_{12_7},r_{12_2}
Jul.1993-Jun.1994	2434	0.4	11	-10.6	0.01	r_{2_1}
Jul.1994-Jun.1995	2548	2.4	11	-8.6	0.01	r_{2_1}
Jul.1995-Jun.1996	2741	14.1	22	-7.9	0.02	BEME,r_{2_1}
Jul.1996-Jun.1997	2928	18.1	22	-3.9	0.01	BEME,r_{2_1}
Jul.1997-Jun.1998	2894	26.5	33	-6.5	0.02	r_{2_1},r_{12_7},r_{12_2}
Jul.1998-Jun.1999	2905	24.6	33	-8.4	0.02	AT,LME,r_{2_1}
Jul.1999-Jun.2000	2804	13.8	22	-8.2	0.03	r_{2_1},r_{12_7}
Jul.2000-Jun.2001	2570	37.7	44	-6.3	0.02	AT,LME, r_{2_1}, r_{6_2}
Jul.2001-Jun.2002	2516	1.3	11	-9.7	0.02	r_{2_1}
Jul.2002-Jun.2003	2491	15	22	-7	0.02	Lev, r_{2_1}
Jul.2003-Jun.2004	2402	3.9	11	-7.1	0.01	r_{2_1}
Jul.2004-Jun.2005	2326	1.8	11	-9.2	0.01	IPM
Jul.2005-Jun.2006	2241	2.5	11	-8.5	0.01	r_{2_1}
Jul.2006-Jun.2007	2178	1.5	11	-9.5	0.01	r_{2_1}
Jul.2007-Jun.2008	2113	12.6	20	-7.4	0.01	r_{12_2},r_{2_1}
Jul.2008-Jun.2009	2023	1.7	10	-8.3	0.02	r_{2_1}
Jul.2009-Jun.2010	2007	1	10	-9	0.01	r_{2_1}
Jul.2010-Jun.2011	1924	13.6	20	-6.4	0.01	r_{2_1}
Jul.2011-Jun.2012	1990	2.5	10	-7.5	0.01	r_{2_1}
Jul.2012-Jun.2013	1937	23.7	30	-6.3	0.01	r_{2_1},r_{12_7},r_{12_2}
Jul.2013-Jun.2014	1909	2.3	10	-7.7	0.01	r_{2_1}
Jul.2014-Jun.2015	1872	5.5	10	-4.5	0.01	r_{2_1}
Jul.2015-Jun.2016	1841	12.4	20	-7.6	0.01	DelGmSale,r_{2_1}
Jul.2016-Jun.2017	1826	26.1	30	-3.9	0.01	C2D,PCM,r_{12_7}

This table summarizes the empirical results, where n represents the number of stock in this rolling window.

3.8.5 Dynamic Peer Groups of Arbitrage Characteristics

In this section, we illustrate that there are distinguishable peer groups of the same arbitrage characteristic resulting in similar mispricing returns. We apply the methods in section 3.5 and take two rolling windows, namely, Jul.1986- Jun.1987 and Jul.2004-Jun.2005 as demonstrative examples.

In the rolling window Jul.1986-Jun.1987, PCM is selected as the only arbitrage characteristic that can explain arbitrage returns. We reveal that similar characteristic-based arbitrage returns are determined by distinguishable groups of the characteristic PCM. We first divide arbitrage returns $\ddot{y}_{it}$ into different return groups. And then, we detect whether there are some clustering structures within groups of the highest and the lowest characteristic-based arbitrage returns, respectively. As we have 2326 assets, for the visualization purpose, we set the threshold value of the K-means method to be relatively small to have as many as ten groups.

Table 3.6 First layer 1986-1987 (clusterings of $\ddot{y}_{it}$)

Group number	Group centroid	Group size
1	0.0059	435
2	0.1205	26
3	-0.0082	428
4	0.0399	189
5	0.0697	71
6	-0.1018	29
7	-0.0617	110
8	-0.0390	250
9	-0.0225	349
10	0.0208	386

In Table 3.6, group 2 has the largest positive average return while group 6 has the worst. Next, we detect the clusterings of characteristic "PCM" within each group individually, which is the second layer in section 3.5.

Table 3.7 Second layer 1986-1987 (clusterings of characteristic PCM)

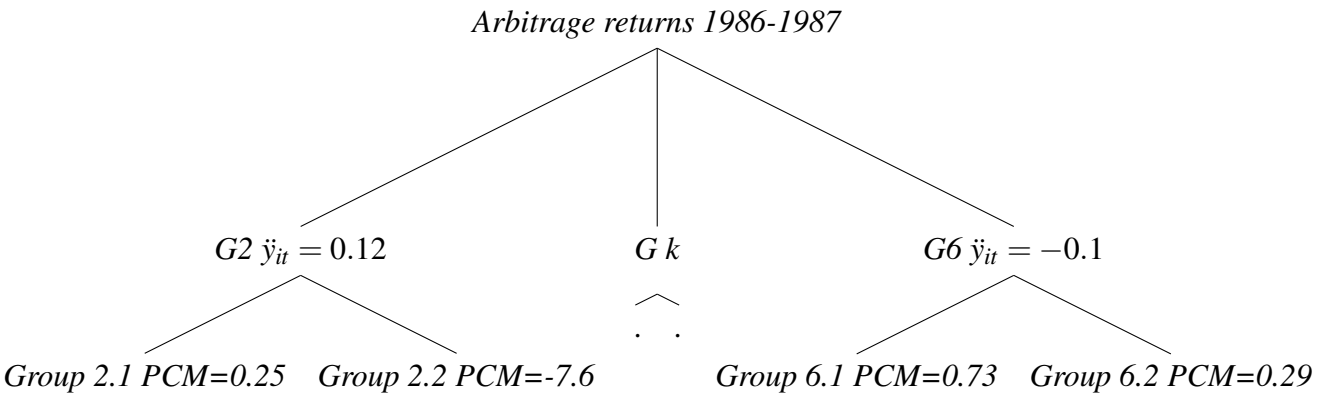
Group number	Centroids of Arbitrage returns	Centroids of PCM	Group size
2.1	0.1211	0.2452	25
2.2	0.1039	-7.630	1

In Table 3.7, there are two clusterings of PCM that provide the highest positive characteristic-based arbitrage returns. Group 2.2, which has an extreme negative PCM value but a high characteristic-based arbitrage return, is an outlier. Members in group 2.1 with excellent arbitrage performance have positive and small PCM values.

Table 3.8 Second layer 1986-1987 (clusterings of characteristic PCM)

Group number	Centroids of Arbitrage returns	Centroids of PCM	Group size
6.1	-0.1085	0.728	9
6.2	-0.0989	0.288	20

Table 3.8 gives groups of PCM in group 6. Members of this group are divided into two clusterings. Group 6.1 has a relatively large PCM value, while group 6.2 has a smaller PCM, which is close to that in group 2.1 with the highest arbitrage return. This is an evident illustration of the nonlinear structure of $h(\boldsymbol{X})$ in this window. The structure of characteristic-based arbitrage returns during Jul.1986- Jun.1987 is:



The classification can be found at Figure 11, where assets are labeled by their "PERMNO," and both axes are rescaled.

Another example is the characteristic-based arbitrage return $\ddot{y}_{it}$ during the year 2004-2005. The power-enhanced test selects characteristic "IPM" as the only explanatory variable.

We apply the Hierarchical K-means method. The results of the first layer classification can be found in Table 3.9. There are ten groups in total according to the similarity of characteristic-based arbitrage returns. Next, we pick two groups with the highest and the lowest returns, respectively, to give clusterings of "IPM" in these two groups.

Table 3.9 First layer 2004-2005 (clusterings of $\ddot{y}_{it}$)

Group number	Group centroid	Group size
1	0.0421	276
2	0.0059	459
3	0.1537	26
4	-0.024	367
5	0.0659	166
6	0.023	387
7	0.0999	120
8	-0.0758	67
9	-0.0437	244
10	-0.0082	436

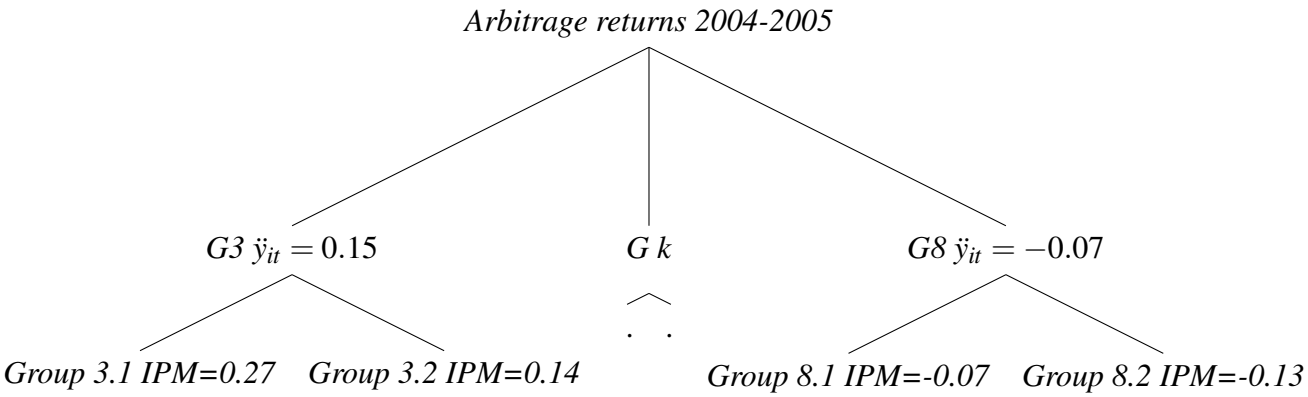
Similarly, we show classification results in Table 3.10 and Table 3.11. Positive IPM values give higher characteristic-based arbitrage returns. On the contrary, when IPM is close to zero or negative, the characteristic-based arbitrage returns fall into the lowest group (group 8).

Table 3.10 Second layer 2004-2005 (clusterings of characteristic IPM)

Group number	Centroids of Arbitrage returns	Centroids of PCM	Group size
3.1	0.1681	0.266	5
3.2	0.1502	0.143	21

Table 3.11 Second layer 2004-2005 (clusterings of characteristic IPM)

Group number	Centroids of Arbitrage returns	Centroids of PCM	Group size
8.1	-0.0713	-0.07	10
8.2	-0.1016	-0.134	57



The plots of the IPM can be found at Figure 12, where the axes are rescaled, and assets are labeled by their "PERMNO" code with five digits.

Finally, it is obvious that peer groups of arbitrage characteristics are dynamic in two aspects. Firstly, the selected arbitrage characteristics are time-varying. Although some of the arbitrage characteristics can show up for more than one block once they appeared, no arbitrage characteristic can be substantially persistent. Secondly, as in Figure 4, the same arbitrage characteristic can have different functional forms in various rolling windows. However, the patterns of some characteristics show persistence in different time periods, such as AT in Figure 6 and LME in Figure 5. In a word, under the flexible semiparametric setting, methods for constructing arbitrage portfolio in Kim et al. (2019) are inapplicable, although the characteristics-based mispricing function is significant for certain time periods.

3.9 Conclusion

We proposed a semiparametric characteristics-based factor asset pricing model, with a focus on the existence and the structure of the mispricing function. Both unknown characteristics-based factor loadings and the mispricing component are approximated by B-splines sieve. The model is estimated by both Project-PCA and equality-constrained OLS. We also develop a power-enhanced test to investigate whether there are mispricing characteristics, orthogonal to the main systematic factors. This is necessary because when the B-splines coefficients of the mispricing function are diverging, the conventional Wald test has very low power. The traditional Wald test is strengthened by a screening process, which adds significant components to the original statistics. Our proposed methods work well for both simulations and the US stock market. Empirically, we find distinct clusterings of the same characteristics resulting in similar arbitrage returns, forming "peer groups." The mispricing functions are time-varying and mostly insignificant under our setting.

Appendices

(A) Characteristic Description

Table 12 Characteristic Details

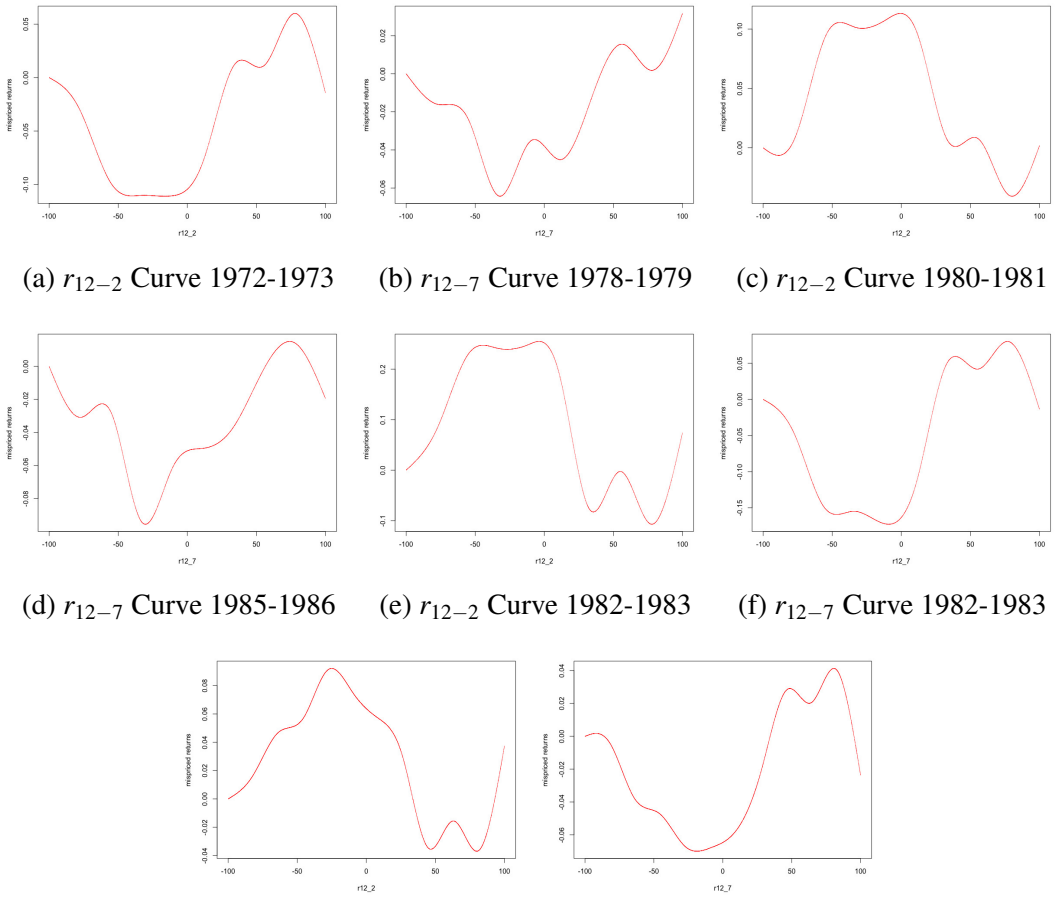
Name	Description	Reference
A2ME	We define assets-market cap as total assets (AT) over market capitalization as of December t-1. Market capitalization is the product of shares outstanding (SHROUT) and price(PRC).	Bhandari (1988)
AT	Total assets (AT)	Gandhi and Lusting (2015)
ATO	Net sales over lagged net operating assets. Net operating assets are the difference between operating assets and operating liabilities. Operating assets are total assets (AT) minus cash and short-term investments (CHE), minus investment and other advances (IVAO). Operating liabilities are total assets (AT), minus debt in current liabilities(DLC),minus long-term debt (DLTT),minus minority interest (MIB), minus preferred stock (PSTK), minus common equity (CEQ).	Soliman(2008)
BEME	Ratio of book value of equity to market value of equity. Book equity is shareholder equity (SH) plus deferred taxes and investment tax credit (TXDITC), minus preferred stock (PS). SH is shareholder‘s equity (SEQ). If missing, SH is the sum of common equity (CEQ) and preferred stock (PS). If missing, SH is the difference between total assets (AT) and total liabilities (LT). Depending on availability, we use the redemption (item PSTKRV), liquidating (item PSTKL), or par value (item PSTK) for PS. The market value of equity is as of December t-1. The market value of equity is the product of shares outstanding (SHROUT) and price (PRC).	Rosenberg, Reid and Lanstein (1985) Davis, Fama, and French (2000)
C	Ration of cash and short-term investments (CHE) to total assets (AT)	Palazzo

C2D	Cash flow to price is the ratio of income and extraoridinary items (IB) and depreciation and amortization (dp) to total liabilities (LT).	
CTO	We define caoital turnover as ratio of net sales (SALE) to lagged total assets (AT).	Haugen and Baker (1996)
Debt2P	Debt to price is the radio of long-term debt (DLTT) and debt in current liabilities (DLC) to the market capitalization as of December t-1 . Market capitalization is the product of shares outstanding (SHROUT) and price (PRC).	Litzenberger and Ramaswamy (1979)
Δceq	The percentage change in the book value of equity (CEQ).	Richardson et al. (2005)
$\Delta(\Delta Gm - Sales)$	The difference in the percentage change in gross margin and the percentage change in sales (SALE). We define gross margin as the difference in sales (SALE) and costs of goods sold (COGS).	Abarbanell and Bushee (1997)
$\Delta Shrout$	The definition of the percentage change in shares outstanding (SHROUT).	Pontiff and Woodgate (2008)
$\Delta PI2A$	We define the change in property, plants ,and equipment as changes in property,plants,and equipment (PPEGT) and inventory (INVT) over lagged total assets (TA).	Lyandres , Sun, and Zhang (2008)
DTO	We define turnover as ratio of daily volume (VOL) to shares outstanding (SHROUT) minus the daily market turnover and de-trend it by its 180 trading day median. We scale down the volume of NASDAQ securities by 38% after 1997 and by 50% before that to address the issue of double-counting of volume for NASDAQ securities.	Garfinkel (2009); Anderson and Dyl (2005)
E2P	We define earnings to price as the ratio of income before extraordinary items (IB) to the market capitalization as December t-1 Market capitalization is the product of share outstanding (SHROUT) and price (PRC).	Basu (1983)

EPS	We define earnings per share as the ratio of income before extraordinary items (IB) to share outstanding (SHROUT) as of December t-1	Basu (1997)
Investment	We define investment as the percentage year-on-year growth rate in total assets (AT).	Cooper, Gulen and Schill(2008)
IPM	We define pre-tax profit margin as ratio of pre-tax income (PI) to sales (SALE).	
Lev	leverage is the ratio of long-term debt (DLTT) and debt in the current liabilities (DLC) to the sum of long-term debt, debt in current liabilities, and stockholders' equity (SEQ)	Lewenllen (2015)
LME	Size is the total market capitalization of the previous month defined as price (PRC) times shares outstanding (SHROUT)	Fama and French (1992)
Turnover	Turnover is last month's volume (VOL) over shares outstanding (SHROUT).	Datar, Naik and Radcliffe (1998)
OL	Operating leverage is the sum of cost of goods sold (COGS) and selling, general, and administrative expenses (XSGA) over total assets.	Novy-Marx (2011)
PCM	The price-to-cost margin is the difference between net sales (SALE) and costs of goods sold (COGS) divided by net sales (SALE).	Gorodnichenko and Weber (2016) and D'Acunto, Liu, Pflucger and Wcber (2017)
PM	The profit margin is operating income after depreciation (OIADP) over sales (SALE)	Soliman (2008)
Q	Tobin's Q is total assets (AT), the market value of equity (SHROUT times PRC) minus cash and short-term investments (CEQ) minus deferred taxes (TXDB) scaled by total assets (AT).	
ROA	Return-on-assets is income before extraordinary items (IB) to lagged total assets (AT).	Balakrishnan, Bartov and Faurel (2010)
ROC	ROC is the ratio of market value of equity (ME) plus long-term debt (DLTT)minus total assets to Cash and Short-Term Investments (CHE).	Chandrashekar and Rao (2009)
ROE	Return-on-equity is income before extraordinary items (IB) to lagged book-value of equity.	in Haugen and Baker (1996)

r_{12-2}	We define momentum as cumulative return from 12 months before the return prediction to two months before.	Fama and French (1996)
r_{12-7}	We define intermediate momentum as cumulative return from 12 months before the return prediction to seven months before.	Novy-Marx (2012)
r_{6-2}	We definer_{6-2} as cumulative return from 6 months before the return prediction to two months before.	Jegadeesh and Titman (1993)
r_{2-1}	We define short-term reversal as lagged one-month return.	Jegadeesh(1990)
S2C	Sales-to-cash is the ratio of net sales (SALE) to Cash and Short-Term Investments (CHE).	following Ou and Penman (1989)
Sales-G	Sales growth is the percentage growth rate in annual sales (SALE).	Lakonishok, Shleifer , and Vishmy (1994)
SAT	We define asset turnover as the ratio of sales (SALE) to total assets (AT).	Soliman (2008)
SGA2S	SGA to sales is the ratio of selling ,general and administrative expenses (XSGA) to net sales (SALE).	

(B) Figures

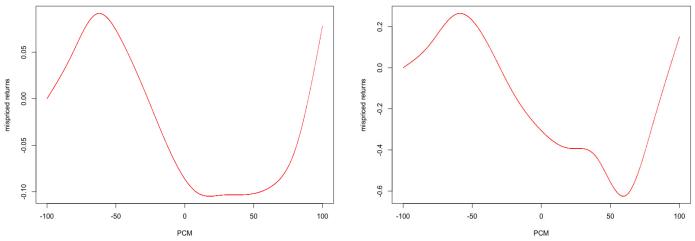


(a) r_{12-2} Curve 1972-1973 (b) r_{12-7} Curve 1978-1979 (c) r_{12-2} Curve 1980-1981

(d) r_{12-7} Curve 1985-1986 (e) r_{12-2} Curve 1982-1983 (f) r_{12-7} Curve 1982-1983

(g) r_{12-2} Curve 1985-1986 (h) r_{12-2} Curve 1985-1986

Figure 1 Mispricing Characteristic Curve of standardized r_{12-2} and r_{12-7}



(a) PCM Curve 1984-1985 (b) PCM Curve 2016-2017

Figure 2 Mispricing Characteristic Curve of standardized PCM

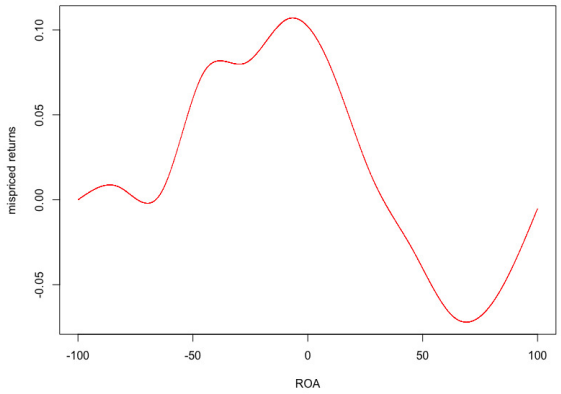
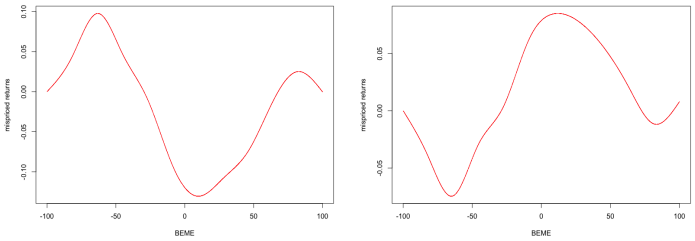
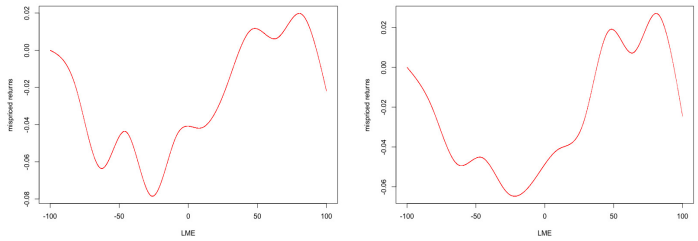


Figure 3 Mispricing Characteristic Curve of standardized ROA in 1988-1989



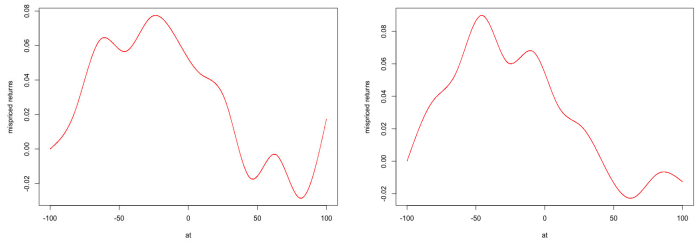
(a) BEME Curve 1995-1996 (b) BEME Curve 1996-1997

Figure 4 Mispricing Characteristic Curve of standardized BEME



(a) LME Curve 1998-1999 (b) LME Curve 2000-2001

Figure 5 Mispricing Characteristic Curve of standardized LME



(a) AT Curve 1998-1999 (b) AT Curve 2000-2001

Figure 6 Mispricing Characteristic Curve of standardized AT

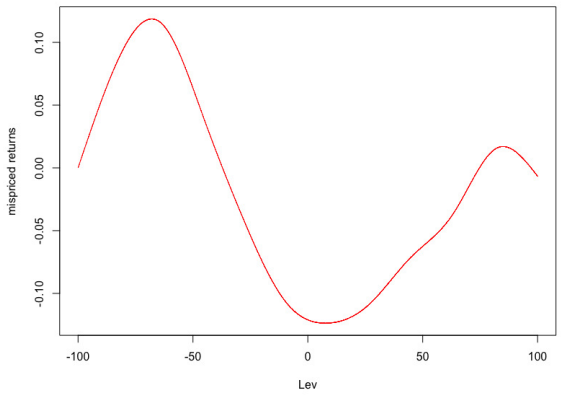


Figure 7 Mispricing Characteristic Curve of standardized LEV in 2002-2003

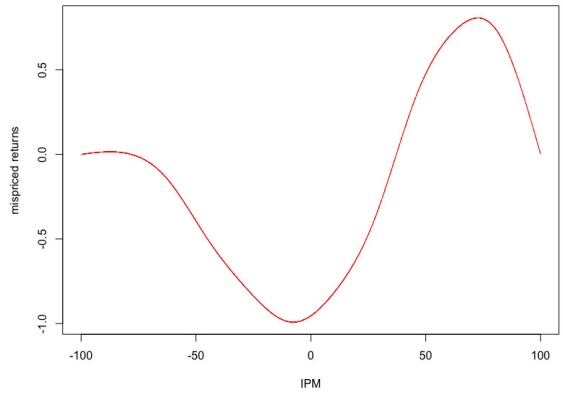


Figure 8 Mispricing Characteristic Curve of standardized IPM in 2004-2005

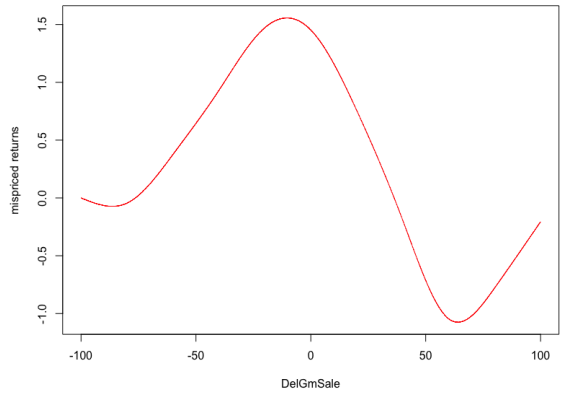


Figure 9 Mispricing Characteristic Curve of standardized DelGmSale in 2015-2016

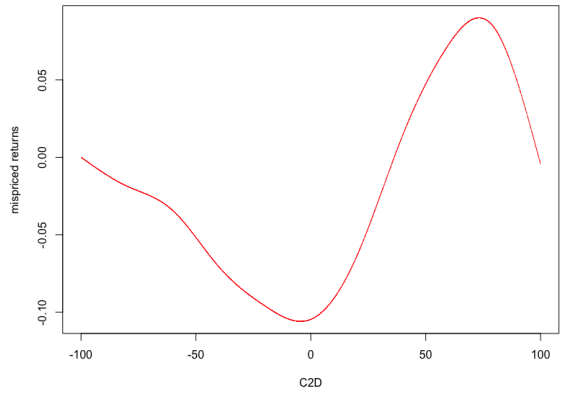
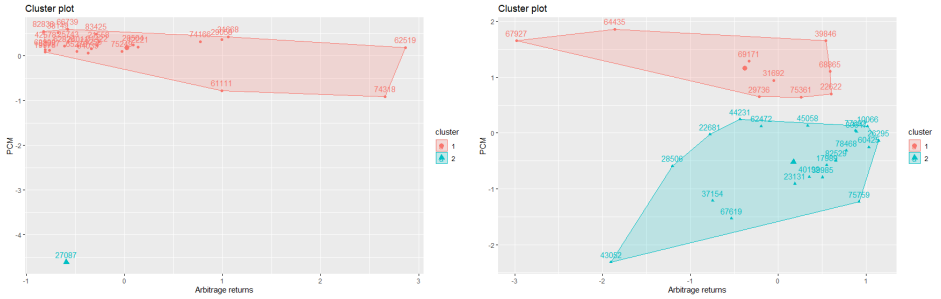


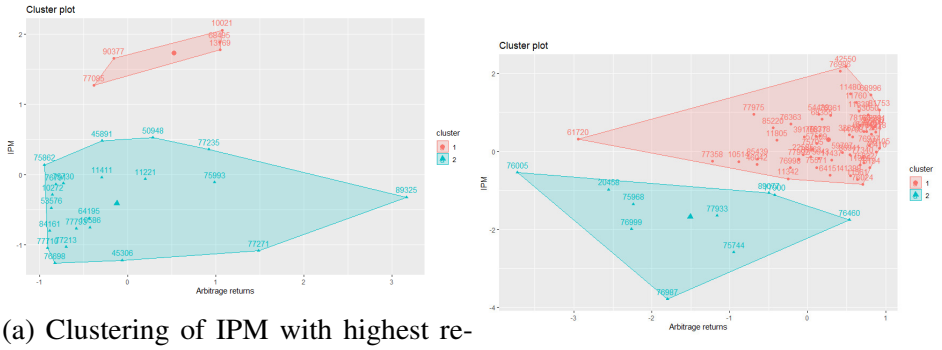
Figure 10 Mispricing Characteristic Curve of standardized C2D in 2016-2017



(a) Clustering of PCM with highest returns

(b) Clustering of PCM with lowest returns

Figure 11 Clustering of PCM 1986-1987



(a) Clustering of IPM with highest returns

(b) Clustering of IPM with lowest returns

Figure 12 Clustering of IPM 2004-2005

(C) Proofs

Throughout the proofs, we have the number of observations $n \to \infty$, and time T is fixed.

Proof of Theorem 4 : In equation 5, we have

$$\boldsymbol{Y} = (\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A} + \boldsymbol{\Gamma} + \boldsymbol{R}^{\mu}(\boldsymbol{X}))\mathbf{1}_T^{\intercal} + (\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B} + \boldsymbol{\Lambda} + \boldsymbol{R}^{\theta}(\boldsymbol{X}))\boldsymbol{F}^{\intercal} + \boldsymbol{U},$$

Multiply time-demeaned matrix $\boldsymbol{D}_T$ on both sides, where $\boldsymbol{D}_T = \boldsymbol{I}_T - \frac{1}{T}\mathbf{1}_T^{\intercal}\mathbf{1}_T$. Given time-invariant mispricing components, we obtain:

$$\boldsymbol{Y}\boldsymbol{D}_T = (\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B} + \boldsymbol{\Lambda} + \boldsymbol{R}^{\theta}(\boldsymbol{X}))\boldsymbol{F}^{\intercal}\boldsymbol{D}_T + \boldsymbol{U}\boldsymbol{D}_T.$$

On-wards, we define $\boldsymbol{Y}\boldsymbol{D}_T = \tilde{\boldsymbol{Y}}$ and $\boldsymbol{F}^{\intercal} = \boldsymbol{F}^{\intercal}\boldsymbol{D}_T$. Time-demeaned factors do not change their properties.

Next, multiple both sides by $\boldsymbol{P} = \boldsymbol{\Phi}(\boldsymbol{X})(\boldsymbol{\Phi}(\boldsymbol{X})^{\intercal}\boldsymbol{\Phi}(\boldsymbol{X}))^{-1}\boldsymbol{\Phi}(\boldsymbol{X})^{\intercal}$,

$$\hat{\boldsymbol{Y}} = (\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B} + \boldsymbol{P}\boldsymbol{\Lambda} + \boldsymbol{P}\boldsymbol{R}^{\theta}(\boldsymbol{X}))\boldsymbol{F}^{\intercal} + \boldsymbol{P}\boldsymbol{U}\boldsymbol{D}_T.$$

We decompose:

$$\boldsymbol{P}\tilde{\boldsymbol{Y}} = \hat{\boldsymbol{Y}} = \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B}\boldsymbol{F}^{\intercal} + \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{F}^{\intercal} + \boldsymbol{P}\boldsymbol{U}\boldsymbol{D}_T + \boldsymbol{P}\boldsymbol{R}^{\theta}(\boldsymbol{X})\boldsymbol{F}^{\intercal} = \boldsymbol{e}_1 + \boldsymbol{e}_2 + \boldsymbol{e}_3 + \boldsymbol{e}_4,$$

as $n \to \infty$ and $n^v \to \infty$, approximation error $\boldsymbol{R}^{\theta}(\boldsymbol{X}) \to_P \mathbf{0}$ as in Huang et al. (2010). Thus, $\boldsymbol{e}_4^{\intercal} \to^P \mathbf{0}$.

Under Assumption 10, we have following results:

for $\frac{1}{n}\sum_{j=1}^{3}\boldsymbol{e}_2^{\intercal}\boldsymbol{e}_j$,

$$\frac{1}{n}\boldsymbol{P}\boldsymbol{\Lambda} \to^P \mathbf{0},$$

therefore,

$$\frac{1}{n}\sum_{j=1}^{3}\boldsymbol{e}_2^{\intercal}\boldsymbol{e}_j + \frac{1}{n}\sum_{j=1}^{3}\boldsymbol{e}_j^{\intercal}\boldsymbol{e}_2 \to^P \mathbf{0}.$$

For $\frac{1}{n}\sum_{j=1}^{3}\boldsymbol{e}_3^{\intercal}\boldsymbol{e}_j$,

$$\frac{1}{n}\boldsymbol{PU} \to^{P} \mathbf{0},$$

therefore,

$$\frac{1}{n}\sum_{j=1}^{3}\boldsymbol{e}_2^{\intercal}\boldsymbol{e}_j + \frac{1}{n}\sum_{j=1}^{3}\boldsymbol{e}_j^{\intercal}\boldsymbol{e}_2 \to^{P} \mathbf{0}.$$

And only $\frac{1}{n}\boldsymbol{e}_1^{\intercal}\boldsymbol{e}_1$ left, namely,

$$\frac{1}{n}\boldsymbol{e}_1^{\intercal}\boldsymbol{e}_1 = \boldsymbol{F}\frac{B^{\intercal}\Phi^{\intercal}(\boldsymbol{X})\Phi(\boldsymbol{X})B}{n}\boldsymbol{F}^{\intercal}.$$

Under Assumption 11-13 and fixed T. A much smaller $T \times T$ matrix $\frac{1}{n}\hat{\boldsymbol{Y}}^{\intercal}\hat{\boldsymbol{Y}}$ can be sovled by asymptotic principal component by Connor and Korajczyk (1986). $\hat{\boldsymbol{F}} = \frac{1}{\sqrt{T}}\{\psi_1, \psi_2, \ldots, \psi_J\}$, where$\{\psi_1, \psi_2, \ldots, \psi_J\}$ are eigenvectors corresponding to the first J eigenvalues of $\frac{1}{n}\hat{\boldsymbol{Y}}^{\intercal}\hat{\boldsymbol{Y}}$.

Thus, $\hat{\boldsymbol{F}} \to_P \boldsymbol{F}$ follows. □

Proof of Theorem 5 : Given $\hat{\boldsymbol{F}}$, we have:

$$\hat{\boldsymbol{G}}(\boldsymbol{X}) = \hat{\boldsymbol{Y}}\hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}^{\intercal}\hat{\boldsymbol{F}})^{-1},$$

as $\hat{\boldsymbol{F}}^{\intercal}\hat{\boldsymbol{F}} = \boldsymbol{I}_J$, therefore,

$$\hat{\boldsymbol{G}}(\boldsymbol{X}) = \tilde{\boldsymbol{Y}}\hat{\boldsymbol{F}}.$$

Then we need to show:

$$E((\hat{\boldsymbol{G}}(\boldsymbol{X_i}) - \boldsymbol{G}(\boldsymbol{X_i}))^2) = 0.$$

Take the sample analogue,

$$\frac{1}{n}(\hat{\boldsymbol{G}}(\boldsymbol{X}) - \boldsymbol{G}(\boldsymbol{X}))^{\intercal}(\hat{\boldsymbol{G}}(\boldsymbol{X}) - \boldsymbol{G}(\boldsymbol{X})).$$

Given:

$$\boldsymbol{G}(\boldsymbol{X}) = \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B} + \boldsymbol{R}^{\theta}(\boldsymbol{X}).$$

$$\hat{\boldsymbol{G}}(\boldsymbol{X}) = (\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B} + \boldsymbol{P}\boldsymbol{\Lambda} + \boldsymbol{P}\boldsymbol{R}^{\boldsymbol{\theta}}(\boldsymbol{X}))\boldsymbol{F}^{\intercal}\hat{\boldsymbol{F}} + \boldsymbol{P}\boldsymbol{U}\boldsymbol{D}_T\hat{\boldsymbol{F}}$$

Furthermore,

$$\boldsymbol{G}(\boldsymbol{X}) - \hat{\boldsymbol{G}}(\boldsymbol{X}) = (\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B} + \boldsymbol{P}\boldsymbol{\Lambda} + \boldsymbol{P}\boldsymbol{R}^{\theta}(\boldsymbol{X}))\boldsymbol{F}^{\intercal}\hat{\boldsymbol{F}} + \boldsymbol{P}\boldsymbol{U}\boldsymbol{D}_T\hat{\boldsymbol{F}} - \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B} - \boldsymbol{R}^{\theta}(\boldsymbol{X}) = \boldsymbol{q}_1 + \boldsymbol{q}_2 + \boldsymbol{q}_3 + \boldsymbol{q}_4.$$

Similar to the Proof of Theorem 4,

$$\frac{1}{n}(\hat{\boldsymbol{G}}(\boldsymbol{X})-\boldsymbol{G}(\boldsymbol{X}))^{\intercal}(\hat{\boldsymbol{G}}(\boldsymbol{X})-\boldsymbol{G}(\boldsymbol{X})) \to^{P} \frac{1}{n}\boldsymbol{q}_1^{\intercal}\boldsymbol{q}_1+\frac{1}{n}\boldsymbol{q}_3^{\intercal}\boldsymbol{q}_3+\frac{1}{n}\boldsymbol{q}_1^{\intercal}\boldsymbol{q}_3+\frac{1}{n}\boldsymbol{q}_3^{\intercal}\boldsymbol{q}_1.$$

For the first term,

$$\frac{1}{n}\boldsymbol{q}_1^{\intercal}\boldsymbol{q}_1=\hat{\boldsymbol{F}}^{\intercal}\boldsymbol{F}(\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B}+\boldsymbol{P}\boldsymbol{\Lambda}+\boldsymbol{P}\boldsymbol{R}^{\theta}(\boldsymbol{X}))^{\intercal}(\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B}+\boldsymbol{P}\boldsymbol{\Lambda}+\boldsymbol{P}\boldsymbol{R}^{\theta}(\boldsymbol{X}))\boldsymbol{F}^{\intercal}\hat{\boldsymbol{F}},$$

due to

$$\frac{1}{n}\sum_{j=1}^{3}\boldsymbol{e}_2^{\intercal}\boldsymbol{e}_j+\frac{1}{n}\sum_{j=1}^{3}\boldsymbol{e}_j^{\intercal}\boldsymbol{e}_2 \to^{P} \boldsymbol{0},$$

and

$$\frac{1}{n}\boldsymbol{e}_1^{T}\boldsymbol{e}_1 \to^{P} \boldsymbol{F}\frac{\boldsymbol{B}^{\intercal}\boldsymbol{\Phi}^{\intercal}(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B}}{n}\boldsymbol{F}^{\intercal}$$

then,

$$\frac{1}{n}\boldsymbol{q}_1^{T}\boldsymbol{q}_1 \to^{P} \hat{\boldsymbol{F}}^{\intercal}\boldsymbol{F}\frac{\boldsymbol{B}^{\intercal}\boldsymbol{\Phi}^{\intercal}(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B}}{n}\boldsymbol{F}^{\intercal}\hat{\boldsymbol{F}}.$$

Theorem 4 and Assumption 11 give $\hat{\boldsymbol{F}} \to \boldsymbol{F}$ and $\boldsymbol{F}^{T}\boldsymbol{F}=\boldsymbol{I}_J$, therefore:

$$\frac{1}{n}\boldsymbol{q}_1^{T}\boldsymbol{q}_1 \to^{P} \frac{\boldsymbol{B}^{\intercal}\boldsymbol{\Phi}^{\intercal}(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B}}{n},$$

Similarly,

$$\frac{1}{n}\boldsymbol{q}_3^{T}\boldsymbol{q}_3 \to^{P} \frac{\boldsymbol{B}^{\intercal}\boldsymbol{\Phi}^{\intercal}(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B}}{n},$$

$$\frac{1}{n}\boldsymbol{q}_1^{T}\boldsymbol{q}_3 \to^{P} -\frac{\boldsymbol{B}^{\intercal}\boldsymbol{\Phi}^{\intercal}(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B}}{n},$$

$$\frac{1}{n}\boldsymbol{q}_3^{T}\boldsymbol{q}_1 \to^{P} -\frac{\boldsymbol{B}^{\intercal}\boldsymbol{\Phi}^{\intercal}(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B}}{n}.$$

Therefore,

$$\frac{1}{n}\boldsymbol{q}_1^{\intercal}\boldsymbol{q}_1+\frac{1}{n}\boldsymbol{q}_3^{\intercal}\boldsymbol{q}_3+\frac{1}{n}\boldsymbol{q}_1^{\intercal}\boldsymbol{q}_3+\frac{1}{n}\boldsymbol{q}_3^{\intercal}\boldsymbol{q}_1 \to 0.$$

Then,

$$\frac{1}{n}(\hat{\boldsymbol{G}}(\boldsymbol{X})-\boldsymbol{G}(\boldsymbol{X}))^{\intercal}(\hat{\boldsymbol{G}}(\boldsymbol{X})-\boldsymbol{G}(\boldsymbol{X})) \to^{P} 0,$$

thus,

$$\hat{\boldsymbol{G}}(\boldsymbol{X}) \to^{P} \boldsymbol{G}(\boldsymbol{X}).$$

Then Theorem 5 follows.

□

Proof of Theorem 6 : Let $\dot{\boldsymbol{Y}} = \frac{1}{T}(\boldsymbol{Y} - \hat{\boldsymbol{G}}(\boldsymbol{X})\hat{\boldsymbol{F}})\mathbf{1}_T$. By substituting the restriction, we have the Lagrangian equation:

$$\min_{\boldsymbol{A}} (\dot{\boldsymbol{Y}} - \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A})^\top (\dot{\boldsymbol{Y}} - \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}) + \boldsymbol{\lambda}\hat{\boldsymbol{G}}^\top(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A} \tag{9}$$

Then we take the first order condition with respect to $\boldsymbol{A}$ and $\boldsymbol{\lambda}$ separately, and we obtain:

$$\begin{pmatrix} 2\boldsymbol{\Phi}(\boldsymbol{X})^\top\boldsymbol{\Phi}(\boldsymbol{X}) & \boldsymbol{\Phi}(\boldsymbol{X})^\top\hat{\boldsymbol{G}}(\boldsymbol{X}) \\ \hat{\boldsymbol{G}}(\boldsymbol{X})^\top\boldsymbol{\Phi}(\boldsymbol{X})^\top & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{A}} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} 2\boldsymbol{\Phi}(\boldsymbol{X})^\top\dot{\boldsymbol{Y}} \\ \mathbf{0} \end{pmatrix}. \tag{10}$$

Under Assumption 11, the above matrices are invertible, which can be written as:

$$\begin{pmatrix} \hat{\boldsymbol{A}} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} 2\boldsymbol{\Phi}(\boldsymbol{X})^\top\boldsymbol{\Phi}(\boldsymbol{X}) & \boldsymbol{\Phi}(\boldsymbol{X})^\top\hat{\boldsymbol{G}}(\boldsymbol{X}) \\ \hat{\boldsymbol{G}}(\boldsymbol{X})^\top\boldsymbol{\Phi}(\boldsymbol{X})^\top & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} 2\boldsymbol{\Phi}(\boldsymbol{X})^\top\dot{\boldsymbol{Y}} \\ \mathbf{0} \end{pmatrix}. \tag{11}$$

Therefore, we obtain:

$$\hat{\boldsymbol{A}} = \boldsymbol{Q}\tilde{\boldsymbol{A}},$$

where

$$\boldsymbol{Q} = \boldsymbol{I} - (\boldsymbol{\Phi}(\boldsymbol{X})^\top\boldsymbol{\Phi}(\boldsymbol{X}))^{-1}\boldsymbol{\Phi}(\boldsymbol{X})^\top\hat{\boldsymbol{G}}(\boldsymbol{X})(\hat{\boldsymbol{G}}(\boldsymbol{X})^\top\hat{\boldsymbol{G}}(\boldsymbol{X}))^{-1}\hat{\boldsymbol{G}}(\boldsymbol{X})^\top\boldsymbol{\Phi}(\boldsymbol{X}),$$

$$\tilde{\boldsymbol{A}} = \frac{1}{T}(\boldsymbol{\Phi}(\boldsymbol{X})^\top\boldsymbol{\Phi}(\boldsymbol{X}))^{-1}\boldsymbol{\Phi}(\boldsymbol{X})^\top\dot{\boldsymbol{Y}}\mathbf{1}_T.$$

Furthermore, let $\boldsymbol{\Xi} = \boldsymbol{\Phi}(\boldsymbol{X})\hat{\boldsymbol{A}} - \boldsymbol{h}(\boldsymbol{X}) = \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{Q}\tilde{\boldsymbol{A}} - \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A} - \boldsymbol{R}^\mu(\boldsymbol{X})$.

Under the restriction $\hat{\boldsymbol{G}}(\boldsymbol{X})^\top\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A} = \mathbf{0}$, we can obtain:

$$\boldsymbol{\Xi} = \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{M}(\boldsymbol{\Phi}(\boldsymbol{X})^\top\boldsymbol{\Phi}(\boldsymbol{X}))^{-1}\boldsymbol{\Phi}(\boldsymbol{X})^\top\frac{1}{T}(\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A} + \boldsymbol{R}^\mu(\boldsymbol{X}) + \boldsymbol{\Gamma} + (\boldsymbol{\Lambda} + \boldsymbol{R}^\theta(\boldsymbol{X}))\boldsymbol{F}')\mathbf{1}_T - \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A} - \boldsymbol{R}^\mu(\boldsymbol{X}). \tag{12}$$

Furthermore, we have:

$$\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{M}(\boldsymbol{\Phi}(\boldsymbol{X})^\top\boldsymbol{\Phi}(\boldsymbol{X}))^{-1}\boldsymbol{\Phi}(\boldsymbol{X})^\top = (\boldsymbol{I} - \boldsymbol{\Phi}(\boldsymbol{X})(\boldsymbol{\Phi}(\boldsymbol{X})^\top\boldsymbol{\Phi}(\boldsymbol{X}))^{-1}\boldsymbol{\Phi}(\boldsymbol{X})^\top\hat{\boldsymbol{G}}(\boldsymbol{X})(\hat{\boldsymbol{G}}(\boldsymbol{X})^\top\hat{\boldsymbol{G}}(\boldsymbol{X}))^{-1}\hat{\boldsymbol{G}}(\boldsymbol{X})^\top)\boldsymbol{P}. \tag{13}$$

And then, substitute Equation 13 into Equation 12 and under Assumption 10 and Theorem 5:

$$\boldsymbol{\Xi} = \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A} - \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A} - \boldsymbol{R}^\mu(\boldsymbol{X}).$$

$$\boldsymbol{R}^{\mu}(\boldsymbol{X}) \to \mathbf{0} \text{ as } n \to \infty,$$

therefore,

$$\frac{1}{n}\boldsymbol{\Xi}^{\intercal}\boldsymbol{\Xi} \to \mathbf{0}.$$

And the Theorem 6 follows. □

Proof of Theorem 7 : Define $Z = \max_{\{1\leqslant p\leqslant P, 1\leqslant h\leqslant H_n\}} \{|\hat{\alpha}_{ph}|/\hat{\sigma}_{ph}\}$. Under Assumption 3, we have

$$\hat{\alpha}_{ph}/\hat{\sigma}_{ph}|\boldsymbol{H}_0 \to^{d} N(0,1).$$

Therefore, under the $\boldsymbol{H}_0$, we have:

$$\begin{aligned} e^{tE(Z)} &\leqslant E[e^{tZ}] \\ &= E[max\{t|\hat{\alpha}_{ph}|/\hat{\sigma}_{ph}\}] \\ &\leqslant \sum_{p=1,h=1}^{p=P,h=H_n} E[e^{t|\hat{\alpha}_{ph}|/\hat{\sigma}_{ph}}] \\ &= PH_n e^{t^2/2}. \end{aligned}$$

Then take the logarithm of both sides we can obtain:

$$E[Z] \leqslant \frac{\log PH_n}{t} + \frac{t}{2}.$$

If we set $t = \sqrt{2\log PH_n}$ to minimise $\frac{\log PH_n}{t} + \frac{t}{2}$, then we have:

$$E[Z] \leqslant \sqrt{2\log PH_n}.$$

Therefore, we can bound the $|\hat{\alpha}_{ph}|/\hat{\sigma}_{ph}$ by $\sqrt{2\log PH_n}$. □

Proof of Theorem 8 : To proof

$$\Pr(\text{reject } H_0|\hat{\mathscr{M}} \neq \emptyset) \to 1,$$

equivalently, we need to prove

$$\Pr(S_0 + S_1|\hat{\mathscr{M}} \neq \emptyset) \to 1$$

$S_0 = H_n \sum_{p=1}^{P} \boldsymbol{I}(\sum_{h=1}^{H_n} |\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \geqslant \eta_n)$, as $H_n = n^v \to \infty$ when $n \to \infty$.

Once $\hat{\mathscr{M}} \neq \emptyset$, then $\sum_{p=1}^{P} \boldsymbol{I}(\sum_{h=1}^{H_n} |\hat{\alpha}_{ph}|/\hat{\sigma}_{ph} \geqslant \eta_n) \geqslant 1$, therefore, $S_0 \to \infty$ as $n \to \infty$. Meanwhile $F_q = O(1)$, we can show that:

$$\Pr(S_0 + S_1 > F_q | \hat{\mathscr{M}} \neq \emptyset) \to 1.$$

Then the Theorem 8 follows. □

Bibliography

D. Acemoglu, V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi. The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016, 2012.

Y. Ait-Sahalia, R. J. Laeven, and L. Pelizzon. Mutual excitation in eurozone sovereign cds. *Journal of Econometrics*, 183(2):151–167, 2014.

Y. Aït-Sahalia, J. Cacho-Diaz, and R. J. Laeven. Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117(3):585–606, 2015.

J. H. Albert and S. Chib. Bayes inference via gibbs sampling of autoregressive time series subject to markov mean and variance shifts. *Journal of Business & Economic Statistics*, 11(1):1–15, 1993.

U. Ali and D. Hirshleifer. Shared analyst coverage: Unifying momentum spillover effects. *Journal of Financial Economics*, 136(3):649–675, 2020.

M. Aquaro, N. Bailey, and M. H. Pesaran. Estimation and inference for spatial models with heterogeneous coefficients: an application to us house prices. *Journal of Applied Econometrics*, 2020.

I.-M. Baek, A. Bandopadhyaya, and C. Du. Determinants of market-assessed sovereign risk: Economic fundamentals or market risk appetite? *Journal of International Money and Finance*, 24(4):533–548, 2005.

N. Bailey, S. Holly, and M. H. Pesaran. A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. *Journal of Applied Econometrics*, 31 (1):249–280, 2016.

N. Bailey, G. Kapetanios, and M. H. Pesaran. Exponent of cross-sectional dependence for residuals. *Sankhya B*, 81(1):46–102, 2019a.

N. Bailey, M. H. Pesaran, and L. V. Smith. A multiple testing approach to the regularisation of large sample correlation matrices. *Journal of Econometrics*, 208(2):507–534, 2019b.

N. Bailey, G. Kapetanios, and M. H. Pesaran. Measurement of factor strenght: Theory and practice. 2020.

S. R. Baker, N. Bloom, and S. J. Davis. Measuring economic policy uncertainty. *The quarterly journal of economics*, 131(4):1593–1636, 2016.

M. Barigozzi and C. Brownlees. Nets: Network estimation for time series. *Journal of Applied Econometrics*, 34(3):347–364, 2019.

M. Barigozzi and M. Hallin. A network analysis of the volatility of high dimensional financial series. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66 (3):581–605, 2017.

J. Beirne and M. Fratzscher. The pricing of sovereign risk and contagion during the european sovereign debt crisis. *Journal of International Money and Finance*, 34:60–82, 2013.

G. Bekaert and C. R. Harvey. Market integration and contagion. Technical report, National Bureau of Economic Research, 2003.

G. Bekaert, M. Ehrmann, M. Fratzscher, and A. Mehl. The global crisis and equity market contagion. *The Journal of Finance*, 69(6):2597–2649, 2014.

K. Bernoth, J. Von Hagen, and L. Schuknecht. Sovereign risk premiums in the european government bond market. *Journal of International Money and Finance*, 31(5):975–995, 2012.

C. E. Bonferroni. Il calcolo delle assicurazioni su gruppi di teste. *Studi in onore del professore salvatore ortu carboni*, pages 13–60, 1935.

B. H. Boyer, M. S. Gibson, M. Loretan, et al. *Pitfalls in tests for changes in correlations*, volume 597. Board of Governors of the Federal Reserve System Washington, DC, 1997.

M. Caporin, L. Pelizzon, F. Ravazzolo, and R. Rigobon. Measuring sovereign contagion in europe. *Journal of Financial Stability*, 34:150–181, 2018.

M. M. Carhart. On persistence in mutual fund performance. *The Journal of finance*, 52(1): 57–82, 1997.

G. Chamberlain and M. Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica: Journal of the Econometric Society*, pages 1281–1304, 1983.

X. Chen and D. Pouzo. Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica*, 80(1):277–321, 2012.

A. Chudik and M. H. Pesaran. Mean group estimation in presence of weakly cross-correlated estimators. *Economics Letters*, 175:101–105, 2019.

A. Chudik, M. H. Pesaran, and E. Tosetti. Weak and strong cross-section dependence and estimation of large panels, 2011.

L. Cohen and A. Frazzini. Economic links and predictable returns. *The Journal of Finance*, 63(4):1977–2011, 2008.

G. Connor and R. A. Korajczyk. Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of financial economics*, 15(3):373–394, 1986.

G. Connor and O. Linton. Semiparametric estimation of a characteristic-based factor model of common stock returns. *Journal of Empirical Finance*, 14(5):694–717, 2007.

G. Connor, M. Hagmann, and O. Linton. Efficient semiparametric estimation of the fama–french model and extensions. *Econometrica*, 80(2):713–754, 2012.

G. Corsetti, M. Pericoli, and M. Sbracia. 'some contagion, some interdependence': More pitfalls in tests of financial contagion. *Journal of International Money and Finance*, 24(8): 1177–1199, 2005.

D. R. Cox. Note on grouping. *Journal of the American Statistical Association*, 52(280): 543–547, 1957.

R. A. De Santis. The euro area sovereign debt crisis: Identifying flight-to-liquidity and the spillover mechanisms. *Journal of Empirical Finance*, 26:150–170, 2014.

A.-L. Delatte, J. Fouquau, and R. Portes. Regime-dependent sovereign risk pricing during the euro crisis. *Review of Finance*, 21(1):363–385, 2017.

F. X. Diebold and K. Yılmaz. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134, 2014.

E. Dimson. Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics*, 7(2):197–226, 1979.

D. Duffie, L. H. Pedersen, and K. J. Singleton. Modeling sovereign yield spreads: A case study of russian debt. *The journal of finance*, 58(1):119–159, 2003.

A.-M. Dumitru and T. Holden. Quantifying the transmission of european sovereign default risk. 2019.

M. Dungey and V. L. Martin. A multifactor model of exchange rates with unanticipated shocks: measuring contagion in the east asian currency crisis. *Journal of Emerging Market Finance*, 3(3):305–330, 2004.

M. Dungey, R. Fry, B. González-Hermosillo, and V. Martin. The transmission of contagion in developed and developing international bond markets. In *Risk Measurement and Systemic Risk, Proceedings of the Third Joint Central Bank Research Conference*, pages 61–74. Citeseer, 2002.

M. Dungey*, R. Fry, B. González-Hermosillo, and V. L. Martin. Empirical modelling of contagion: a review of methodologies. *Quantitative finance*, 5(1):9–24, 2005.

M. Dungey, R. Fry, B. González-Hermosillo, and V. Martin. Contagion in international bond markets during the russian and the ltcm crises. *Journal of Financial Stability*, 2(1):1–27, 2006.

S. Edwards. Ldc's foreign borrowing and default risk: An empirical investigation, 1983.

S. Edwards. The pricing of bonds and bank loans in international markets: An empirical analysis of developing countries' foreign borrowing. *European Economic Review*, 30(3): 565–589, 1986.

J. P. Elhorst et al. *Spatial econometrics: from cross-sectional data to spatial panels*, volume 479. Springer, 2014.

J. Engelberg, A. Ozoguz, and S. Wang. Know thy neighbor: Industry clusters, information spillovers, and market efficiency. *Journal of Financial and Quantitative Analysis*, 53(5): 1937–1961, 2018.

F. Eugene. The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–465, 1992.

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56, 1993.

E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22, 2015.

J. Fan, Y. Fan, and J. Lv. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1):186–197, 2008.

J. Fan, Y. Liao, and M. Mincheva. High dimensional covariance matrix estimation in approximate factor models. *Annals of statistics*, 39(6):3320, 2011.

J. Fan, Y. Liao, and M. Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 75(4), 2013.

J. Fan, Y. Liao, and J. Yao. Power enhancement in high-dimensional cross-sectional tests. *Econometrica*, 83(4):1497–1541, 2015.

J. Fan, A. Furger, and D. Xiu. Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data. *Journal of Business & Economic Statistics*, 34(4):489–503, 2016a.

J. Fan, Y. Liao, and W. Wang. Projected principal component analysis in factor models. *Annals of statistics*, 44(1):219, 2016b.

C. Favero and A. Missale. Sovereign spreads in the eurozone: which prospects for a eurobond? *Economic Policy*, 27(70):231–273, 2012.

G. Feng, S. Giglio, and D. Xiu. Taming the factor zoo. *Fama-Miller Working Paper*, 24070, 2017.

W. D. Fisher. On grouping for maximum homogeneity. *Journal of the American statistical Association*, 53(284):789–798, 1958.

K. J. Forbes and R. Rigobon. No contagion, only interdependence: measuring stock market comovements. *The journal of Finance*, 57(5):2223–2261, 2002.

J. Freyberger, A. Neuhierl, and M. Weber. Dissecting characteristics nonparametrically. Technical Report 5, 2020.

S. Frühwirth-Schnatter. *Finite mixture and Markov switching models*. Springer Science & Business Media, 2006.

X. Gabaix. The granular origins of aggregate fluctuations. *Econometrica*, 79(3):733–772, 2011.

D. Garcia. Sentiment during recessions. *The Journal of Finance*, 68(3):1267–1300, 2013.

M. R. Gibbons, S. A. Ross, and J. Shanken. A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, pages 1121–1152, 1989.

G. Hale and J. A. Lopez. Monitoring banking system connectedness with big data. *Journal of Econometrics*, 212(1):203–220, 2019.

A. G. Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 33(3):438–443, 1971a.

A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971b.

E. Hjalmarsson and P. Manchev. Characteristic-based mean-variance portfolio choice. *Journal of Banking & Finance*, 36(5):1392–1401, 2012.

G. Hoberg and G. Phillips. Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5):1423–1465, 2016.

K. Hou, C. Xue, and L. Zhang. Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705, 2015.

J. Huang, J. L. Horowitz, and F. Wei. Variable selection in nonparametric additive models. *Annals of statistics*, 38(4):2282, 2010.

J. E. Ingersoll Jr. Some results in the theory of arbitrage pricing. *The Journal of Finance*, 39 (4):1021–1039, 1984.

R. D. Israelsen. Does common analyst coverage explain excess comovement? *Journal of Financial and Quantitative Analysis*, 51(4):1193–1229, 2016.

S. Kaufmann. K-state switching models with time-varying transition distributions—does loan growth signal stronger effects of variables on inflation? *Journal of Econometrics*, 187 (1):82–94, 2015.

M. Kaustia and V. Rantala. Common analyst-based method for defining peer firms. *Available at SSRN*, 2013.

Z. T. Ke, B. T. Kelly, and D. Xiu. Predicting returns with text data. Technical report, National Bureau of Economic Research, 2019.

H. H. Kelejian and I. R. Prucha. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1):99–121, 1998.

H. H. Kelejian and I. R. Prucha. A generalized moments estimator for the autoregressive parameter in a spatial model. *International economic review*, 40(2):509–533, 1999.

B. T. Kelly, S. Pruitt, and Y. Su. Instrumented principal component analysis. *Available at SSRN 2983919*, 2017.

B. T. Kelly, S. Pruitt, and Y. Su. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 2019.

S. Kim, R. A. Korajczyk, and A. Neuhierl. Arbitrage portfolios. *Georgia Tech Scheller College of Business Research Paper*, (18-43), 2019.

A. B. Kock and D. Preinerstorfer. Power in high-dimensional testing problems. *Econometrica*, 87(3):1055–1069, 2019.

S. Kou, X. Peng, and H. Zhong. Asset pricing with spatial interaction. *Management Science*, 64(5):2083–2101, 2018.

G. M. Kuersteiner and I. R. Prucha. Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica*, 88(5):2109–2146, 2020.

O. Ledoit, M. Wolf, et al. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2):1024–1060, 2012.

J. P. LeSage. An introduction to spatial econometrics. *Revue d'économie industrielle*, (123): 19–44, 2008.

C. K. Liew. Inequality constrained least-squares estimation. *Journal of the American Statistical Association*, 71(355):746–751, 1976.

F. A. Longstaff, J. Pan, L. H. Pedersen, and K. J. Singleton. How sovereign is sovereign credit risk? *American Economic Journal: Macroeconomics*, 3(2):75–103, 2011.

N. Metiu. Sovereign risk contagion in the eurozone. *Economics Letters*, 117(1):35–38, 2012.

T. J. Moskowitz and M. Grinblatt. Do industries explain momentum? *The Journal of finance*, 54(4):1249–1290, 1999.

C. A. Parsons, R. Sabbatucci, and S. Titman. Geographic lead-lag effects. *The Review of Financial Studies*, 33(10):4721–4770, 2020.

H. Pesaran et al. General diagnostic tests for cross-sectional dependence in panels. *University of Cambridge, Cambridge Working Papers in Economics*, 435, 2004.

H. H. Pesaran and Y. Shin. Generalized impulse response analysis in linear multivariate models. *Economics letters*, 58(1):17–29, 1998.

M. H. Pesaran and A. Pick. Econometric issues in the analysis of contagion. *Journal of Economic Dynamics and Control*, 31(4):1245–1277, 2007.

M. H. Pesaran and R. Smith. Estimating long-run relationships from dynamic heterogeneous panels. *Journal of econometrics*, 68(1):79–113, 1995.

M. H. Pesaran and T. Yamagata. Testing capm with a large number of assets. In *AFA 2013 San Diego Meetings Paper*, 2012.

M. H. Pesaran and T. Yamagata. Testing for alpha in linear factor pricing models with a large number of securities. 2017.

C. Pirinsky and Q. Wang. Does corporate headquarters location matter for stock returns? *The Journal of Finance*, 61(4):1991–2015, 2006.

D. Pollard. Strong consistency of k-means clustering. *The Annals of Statistics*, pages 135–140, 1981.

D. Pollard et al. A central limit theorem for *k*-means clustering. *The Annals of Probability*, 10(4):919–926, 1982.

R. Rigobon. On the measurement of the international propagation of shocks: is the transmission stable? *Journal of International Economics*, 61(2):261–283, 2003.

S. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3): 341–360, 1976.

A. Scherbina and B. Schlusche. Economic linkages inferred from news stories and the predictability of stock returns. *Available at SSRN 2363436*, 2015.

G. Schwenkler and H. Zheng. The network of firms implied by the news. *Available at SSRN 3320859*, 2019.

W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.

W. Sun, J. Wang, Y. Fang, et al. Regularized k-means clustering of high-dimensional data and its asymptotic consistency. *Electronic Journal of Statistics*, 6:148–167, 2012.

L. Tierney. Markov chains for exploring posterior distributions. *the Annals of Statistics*, pages 1701–1728, 1994.

M. Vogt and O. Linton. Classification of non-parametric regression functions in longitudinal data models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1):5–27, 2017.